Methods in Molecular Biology™

VOLUME 173

# Calcium-Binding Protein Protocols

## Volume II
## Methods and Techniques

Edited by

Hans J. Vogel

HUMANA PRESS

# Calcium-Binding Protein Protocols
# Volume II

# METHODS IN MOLECULAR BIOLOGY™

## *John M. Walker,* Series Editor

207. **Human Antibodies for Cancer Therapy:** *Reviews and Protocols,* edited by *Martin Welschof and Jürgen Krauss, 2002*
206. **Endothelin Protocols,** edited by *Janet J. Maguire and Anthony P. Davenport, 2002*
205. ***E. coli* Gene Expression Protocols,** edited by *Peter E. Vaillancourt, 2002*
204. **Molecular Cytogenetics:** *Methods and Protocols,* edited by *Yao-Shan Fan, 2002*
203. ***In Situ* Detection of DNA Damage:** *Methods and Protocols,* edited by *Vladimir V. Didenko, 2002*
202. **Thyroid Hormone Receptors:** *Methods and Protocols,* edited by *Aria Baniahmad, 2002*
201. **Combinatorial Library Methods and Protocols,** edited by *Lisa B. English, 2002*
200. **DNA Methylation Protocols,** edited by *Ken I. Mills and Bernie H, Ramsahoye, 2002*
199. **Liposome Methods and Protocols,** edited by *Subhash C. Basu and Manju Basu, 2002*
198. **Neural Stem Cells:** *Methods and Protocols,* edited by *Tanja Zigova, Juan R. Sanchez-Ramos, and Paul R. Sanberg, 2002*
197. **Mitochondrial DNA:** *Methods and Protocols,* edited by *William C. Copeland, 2002*
196. **Oxidants and Antioxidants:** *Ultrastructural and Molecular Biology Protocols,* edited by *Donald Armstrong, 2002*
195. **Quantitative Trait Loci:** *Methods and Protocols*, edited by *Nicola J. Camp and Angela Cox, 2002*
194. **Post-translational Modification Reactions,** edited by *Christoph Kannicht, 2002*
193. **RT-PCR Protocols,** edited by *Joseph O'Connell, 2002*
192. **PCR Cloning Protocols, 2nd ed.,** edited by *Bing-Yuan Chen and Harry W. Janes, 2002*
191. **Telomeres and Telomerase:** *Methods and Protocols,* edited by *John A. Double and Michael J. Thompson, 2002*
190. **High Throughput Screening:** *Methods and Protocols,* edited by *William P. Janzen, 2002*
189. **GTPase Protocols:** *The RAS Superfamily,* edited by *Edward J. Manser and Thomas Leung, 2002*
188. **Epithelial Cell Culture Protocols,** edited by *Clare Wise, 2002*
187. **PCR Mutation Detection Protocols,** edited by *Bimal D. M. Theophilus and Ralph Rapley, 2002*
186. **Oxidative Stress and Antioxidant Protocols,** edited by *Donald Armstrong, 2002*
185. **Embryonic Stem Cells:** *Methods and Protocols,* edited by *Kursad Turksen, 2002*
184. **Biostatistical Methods,** edited by *Stephen W. Looney, 2002*
183. **Green Fluorescent Protein:** *Applications and Protocols,* edited by *Barry W. Hicks, 2002*
182. **In Vitro Mutagenesis Protocols, 2nd ed.**, edited by *Jeff Braman, 2002*
181. **Genomic Imprinting:** *Methods and Protocols,* edited by *Andrew Ward, 2002*
180. **Transgenesis Techniques, 2nd ed.:** *Principles and Protocols,* edited by *Alan R. Clarke, 2002*
179. **Gene Probes:** *Principles and Protocols,* edited by *Marilena Aquino de Muro and Ralph Rapley, 2002*
178. **Antibody Phage Display:** *Methods and Protocols,* edited by *Philippa M. O'Brien and Robert Aitken, 2001*
177. **Two-Hybrid Systems:** *Methods and Protocols,* edited by *Paul N. MacDonald, 2001*
176. **Steroid Receptor Methods:** *Protocols and Assays,* edited by *Benjamin A. Lieberman, 2001*
175. **Genomics Protocols**, edited by *Michael P. Starkey and Ramnath Elaswarapu, 2001*
174. **Epstein-Barr Virus Protocols,** edited by *Joanna B. Wilson and Gerhard H. W. May, 2001*
173. **Calcium-Binding Protein Protocols, Volume 2:** *Methods and Techniques,* edited by *Hans J. Vogel, 2001*
172. **Calcium-Binding Protein Protocols, Volume 1:** *Reviews and Case Histories,* edited by *Hans J. Vogel, 2001*
171. **Proteoglycan Protocols,** edited by *Renato V. Iozzo, 2001*
170. **DNA Arrays:** *Methods and Protocols,* edited by *Jang B. Rampal, 2001*
169. **Neurotrophin Protocols,** edited by *Robert A. Rush, 2001*
168. **Protein Structure, Stability, and Folding,** edited by *Kenneth P. Murphy, 2001*
167. **DNA Sequencing Protocols,** *Second Edition,* edited by *Colin A. Graham and Alison J. M. Hill, 2001*
166. **Immunotoxin Methods and Protocols**, edited by *Walter A. Hall, 2001*
165. **SV40 Protocols**, edited by *Leda Raptis, 2001*
164. **Kinesin Protocols,** edited by *Isabelle Vernos, 2001*
163. **Capillary Electrophoresis of Nucleic Acids, Volume 2:** *Practical Applications of Capillary Electrophoresis,* edited by *Keith R. Mitchelson and Jing Cheng, 2001*
162. **Capillary Electrophoresis of Nucleic Acids, Volume 1:** *Introduction to the Capillary Electrophoresis of Nucleic Acids,* edited by *Keith R. Mitchelson and Jing Cheng, 2001*
161. **Cytoskeleton Methods and Protocols**, edited by *Ray H. Gavin, 2001*
160. **Nuclease Methods and Protocols,** edited by *Catherine H. Schein, 2001*
159. **Amino Acid Analysis Protocols,** edited by *Catherine Cooper, Nicole Packer, and Keith Williams, 2001*
158. **Gene Knockoout Protocols,** edited by *Martin J. Tymms and Ismail Kola, 2001*
157. **Mycotoxin Protocols,** edited by *Mary W. Trucksess and Albert E. Pohland, 2001*
156. **Antigen Processing and Presentation Protocols,** edited by *Joyce C. Solheim, 2001*
155. **Adipose Tissue Protocols,** edited by *Gérard Ailhaud, 2000*
154. **Connexin Methods and Protocols,** edited by *Roberto Bruzzone and Christian Giaume, 2001*
153. **Neuropeptide Y Protocols**, edited by *Ambikaipakan Balasubramaniam, 2000*
152. **DNA Repair Protocols:** *Prokaryotic Systems*, edited by *Patrick Vaughan, 2000*
151. **Matrix Metalloproteinase Protocols,** edited by *Ian M. Clark, 2001*
150. **Complement Methods and Protocols,** edited by *B. Paul Morgan, 2000*
149. **The ELISA Guidebook**, edited by *John R. Crowther, 2000*

METHODS IN MOLECULAR BIOLOGY™

# Calcium-Binding Protein Protocols

## *Volume 2: Methods and Techniques*

Edited by

**Hans J. Vogel**

*Department of Biological Sciences, University of Calgary*
*Calgary, AB, Canada*

**Humana Press**  **Totowa, New Jersey**

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

**www.humanapress.com**



This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American Standards Institute) Permanence of Paper for Printed Library Materials.

Cover design by Patricia F. Cleary.
Cover illustration: From Fig. 1A in Chapter 3, Vol. 1 "Crystal Structure of Calpain and Insights into $Ca^{2+}$-Dependent Activation" by Zongchao Jia, Christopher M. Hosfield, Peter L. Davies, and John S. Elce.

Production Editor: Kim Hoather-Potter.

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel: 973-256-1699; Fax: 973-256-8341; E-mail: humana@humanapr.com, or visit our Website at www.humanapress.com



Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Main entry under title: Methods in molecular biology™.

Calcium-binding protein protocols / edited by Hans J. Vogel
p. cm. -- (Methods in molecular biology; v. v. 172-)
Includes bibliographical references and index.
Contents: v. 1. Reviews and case studies.
ISBN 0-89603-689-8 (alk. paper)
1. Calcium-binding proteins--Research--Methodology I. Vogel, Hans J. II. Methods in molecular biology (Clifton, N.J.) ; . v. 172, etc.

QP552.C24 C33 2001
572'.69—dc21 01-063354

# Dedication

This book is dedicated to the memory of Dr. J. David Johnson (Columbus, OH) whose untimely death on January 21, 2000 has deeply shocked all his colleagues and friends. David has made numerous excellent contributions to our understanding of calcium-binding proteins. His insight and enthusiasm will be sadly missed.

***Hans J. Vogel,* PhD**

# Preface

Calcium plays an important role in a wide variety of biological processes. This divalent metal ion can bind to a large number of proteins; by doing so it modifies their biological activity or their stability. Because of its distinct chemical properties calcium is uniquely suited to act as an on–off switch or as a light dimmer of biological activities. The two books entitled *Calcium-Binding Protein Protocols* (Volumes I and II) focus on modern experimental analyses and methodologies for the study of calcium-binding proteins. Both extracellular and intracellular calcium-binding proteins are discussed in detail. However, proteins involved in calcium handling (e.g., calcium pumps and calcium channels), fall outside of the scope of these two volumes. Also, calcium-binding proteins involved in bone deposition will not be discussed, as this specific topic has been addressed previously. The focus of these two books is on studies of the calcium-binding proteins and their behavior in vitro and in vivo. The primary emphasis is on protein chemistry and biophysical methods. Many of the methods described will also be applicable to proteins that do not bind calcium.

*Calcium-Binding Protein Protocols* is divided into three main sections. The section entitled *Introduction and Reviews* provides information on the role of calcium in intracellular secondary messenger activation mechanisms. Moreover, unique aspects of calcium chemistry and the utilization of calcium in dairy proteins, as well as calcium-binding proteins involved in blood clotting, are addressed. The second section entitled *Calcium-Binding Proteins: Case Studies* provides a wealth of information about protein purification and characterization strategies, X-ray crystallography, and other studies that are focused on specific calcium-binding proteins. Together, these two sections comprise Volume I of this series. By introducing the various classes of intra- and extracellular calcium-binding proteins and their modes of action, these two sections set the stage and provide the necessary background for the third section. The final section entitled *Methods and Techniques to Study Calcium-Binding Proteins* makes up Volume II of *Calcium-Binding Protein Protocols.* Here the focus is on the use of a range of modern experimental techniques that can be employed to study the solution structure, stability, dynamics, calcium-binding properties, and biological activity of calcium-binding proteins in general. As well, studies of their ligand-binding properties and their distribution in cells are included. In addition to enzymatic assays and more routine spectroscopic and protein chemistry techniques, particular attention has been paid in the second volume to modern NMR approaches, thermodynamic analyses, kinetic mea-

surements such as surface plasmon resonance, strategies for amino acid sequence alignments, as well as fluorescence methods to study the distribution of calcium and calcium-binding proteins in cells. In preparing their chapters, all the authors have attempted to share the little secrets that are required to successfully apply these methods to related proteins. Together the two volumes of *Calcium-Binding Protein Protocols* provide the reader with a host of experimental methods that can be applied either to uncover new aspects of earlier characterized calcium-binding proteins or to study newly discovered proteins.

As more and more calcium-binding proteins are being uncovered through genome sequencing efforts and protein interaction studies (e.g., affinity chromatography, crosslinking, or yeast two-hybrid systems) the time seemed right to collect all the methods used to characterize these proteins in a book. The methods detailed here should provide the reader with the essential tools for their analysis in terms of structure, dynamics, and function. The hope is that these two volumes will contribute to our understanding of the part of the proteome, which relies on interactions with calcium to carry out its functions.

In closing, I would like to thank Margaret Tew for her invaluable assistance with the editing and organization of these two books. Finally, I would like to thank the authors of the individual chapters, who are all experts in this field, for their cooperation in producing these two volumes in a timely fashion.

***Hans J. Vogel, PhD***

# Contents

## Calcium-Binding Protein Protocols

### *Volume I: Reviews and Case Studies*

# Contributors


Jean-René Alattia • *Division of Molecular and Structural Biology, Ontario Cancer Institute, Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada*

James B. Ames • *Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada*

Stefan Bagby • *Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada*

Peter M. Bayley • *Division of Physical Biochemistry, National Institute for Medical Research, London, UK*

Lawrence J. Berliner • *Department of Chemistry, The Ohio State University, Columbus, OH*

Jonathan Boyd • *Department of Biochemistry and OCMS, University of Oxford, South Parks Road, Oxford, UK*

Richard D. Brokx • *Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada*

Iain D. Campbell • *Department of Biochemistry and OCMS, University of Oxford, South Parks Road, Oxford, UK*

Teresa E. Clarke • *Department of Biological Sciences, University of Calgary, Calgary, AB, Canada*

David C. Corson • *Canadian Institutes of Health Research Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, AB, Canada*

Amanda L. Doherty-Kirby • *Department of Biochemistry, University of Western Ontario, London, ON, Canada*

A. Kristina Downing • *Department of Biochemistry and OCMS, University of Oxford, South Parks Road, Oxford, UK*

Torbjörn Drakenberg • *Department of Physical Chemistry 2, Lund University, Lund, Sweden*

Heinz Fabian • *Max-Delbrück-Center for Molecular Medicine, Berlin, Germany*

Thomas Grundström • *Department of Cell and and Molecular Biology, Umeå University, Umeå, Sweden*

Jacques Haiech • *Pharmacologie et Physico-Chimie des Interactions Cellulaires et Moléculaires, Faculté de Pharmacie, Université Louis Pasteur, Strasbourg, Illkirch, France*

Jaap Heringa • *Division of Mathematical Biology, MRC National Institute for Medical Research, London, UK*

Leslie D. Hicks • *Department of Biochemistry, University of Alberta, Edmonton, AB, Canada*
Steven Howell • *National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK*
Kate Hughes • *Department of Cell and and Molecular Biology, Umeå University, Umeå, Sweden*
Mitsuhiko Ikura • *Division of Medical and Structural Biology, Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada*
J. David Johnson • *Department of Molecular and Cellular Biochemistry, The Ohio State University Medical Center, Columbus, OH*
Karin Julenius • *Department of Physical Chemistry, Lund University, Lund, Sweden*
Cyril M. Kay • *Department of Biochemistry, University of Alberta, Edmonton, AB, Canada*
Marie-Claude Kilhoffer • *Pharmacologie et Physico-Chimie des Interactions Cellulaires et Moléculaires, Faculté de Pharmacie, Université Louis Pasteur, Strasbourg, Illkirch, France*
Ryoji Kobayashi • *Department of Chemistry, Kagawa Medical University, Ikenobe, Miki-cho, Kita-gun, Kagawa, Japan*
Ritsu Kondo • *Department of Medical Biochemistry, The Ohio State University Medical Center, Columbus, OH*
Joanna K. Krueger • *Chemistry Department, University of North Carolina, Charlotte, NC*
Gilles A. Lajoie • *Guelph-Waterloo Centre for Graduate Work in Chemistry and Biochemistry, Department of Chemistry, University of Waterloo, Waterloo, ON, Canada*
Monica X. Li • *Canadian Institutes of Health Research Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, AB, Canada*
Sarah Linse • *Department of Physical Chemistry 2, Lund University, Lund, Sweden*
Maria M. Lopez • *Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, PA*
George I. Makhatadze • *Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, PA*
Tapas K. Mal • *Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada*

Stephen R. Martin • *Division of Physical Biochemistry, National Institute of Medical Research, London, UK*
Anthony Persechini • *Department of Pharmacology and Physiology, University of Rochester Medical Center, Rochester, NY*
Juha Saarikettu • *Department of Cell and Molecular Biology, Umeå University, Umeå, Sweden*
Gary S. Shaw • *Department of Biochemistry, The University of Western Ontario, London, ON, Canada*
Cindy Sutherland • *Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada*
Mark B. Swindells • *Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada*
Brian D. Sykes • *Department of Biochemistry, University of Alberta, Edmonton, AB, Canada*
Richard Thorogate • *Department of Pharmacology and Clinical Pharmacology, St. George's Hospital Medical School, Cranmer Terrace, London, UK*
Svetlana B. Tikunova • *Department of Molecular and Cellular Biochemistry, The Ohio State University Medical Center, Columbus, OH*
Katalin Török • *Department of Pharmacology and Clinical Pharmacology, St. George's Hospital Medical School, Cranmer Terrace, London, UK*
Jill Trewhella • *Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM*
Jacquelyn E. Van Lierop • *Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada*
Hans J. Vogel • *Department of Biological Sciences, University of Calgary, Calgary, AB, Canada*
Michael P. Walsh • *Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada*
Aalim M. Weljie • *Department of Biological Sciences, University of Calgary, Calgary, AB, Canada*
Jörn M. Werner • *Department of Biochemistry and OCMS, University of Oxford, South Parks Road, Oxford, UK*
Kyoko L. Yap • *Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto, ON, Canada*
Michio Yazawa • *Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo, Japan*

# III

# Methods and Techniques to Study Calcium-Binding Proteins

# 1

# Quantitative Analysis of $Ca^{2+}$-Binding by Flow Dialysis

**Michio Yazawa**

## 1. Introduction

$Ca^{2+}$-binding to proteins can be measured directly by equilibrium dialysis ***(1,2)***, the standard method for the direct measurement of the binding of small-ligand molecules by macromolecules. In this method, a semipermeable cellulose bag containing a solution of macromolecules is immersed in the buffer solution containing ligand molecules and is incubated to attain both the chemical and diffusion equilibrium. The method can be improved with the use of two small thin chambers separated by the cellulose membrane, which may reduce the incubation time required to achieve diffusion equilibrium (microdialysis) ***(3)***. Ligand molecules are usually labeled with the radioactive isotopes for quantitative determinations, and ligand molecules bound to the macromolecule in the equilibrium state are determined directly from the difference between the free concentration in the dialysate and the total concentration in the protein solution. Binding of ligand to the protein molecule can be calculated from the known value of the protein concentration, and the ligand bindings at several free concentrations of the ligand are determined from independent experiments to yield a ligand binding curve from which the maximum number of ligand binding and the equilibrium constants are estimated. In this method, the ligand-binding equilibrium, which is usually obtained within less than a second, has to be assessed after attainment of the diffusion equilibrium of ligands across the membrane, which usually takes a much longer time — on the order of several hours.

This major drawback in the equilibrium dialysis method has been overcome by the flow-dialysis method ***(4,5)***. In this method, a sample chamber contain-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel © Humana Press Inc., Totowa, NJ

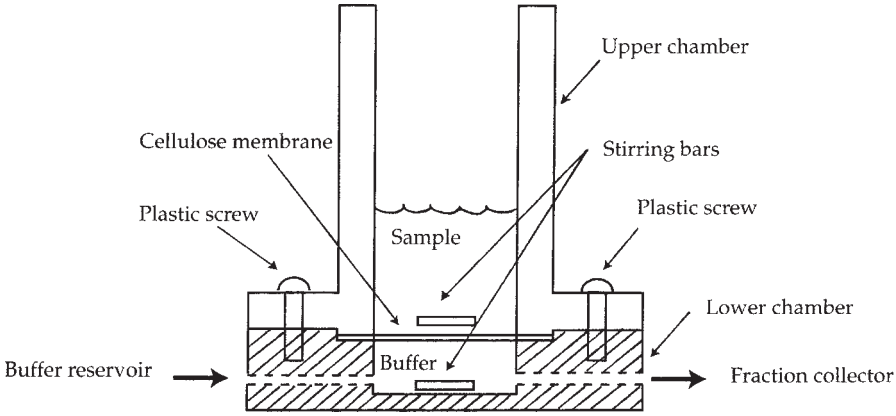


Fig. 1. Schematic representation of a flow-dialysis cell.

ing a protein solution is separated by the cellulose membrane from the buffer chamber filled with the buffer solution (*see* **Fig. 1**). Each solution in both of the chambers is continuously mixed with a magnetic stirring bar, and the concentration of free ligand in the protein solution is determined based on the rate of diffusion into the buffer chamber, which is proportional to the concentration of free ligand. The buffer chamber is connected to the reservoir and is flushed continuously with the fresh buffer solution at a constant rate and the outlet is connected to a fraction collector to monitor the radioactivity in the effluent. When small amounts of the labeled ligand in a small volume are added to the sample chamber, chemical equilibrium is attained usually within a fraction of a second and the ligand molecules free from the protein molecule diffuse into the buffer chamber at a rate depending on the equilibrium concentration and the characteristics of the membrane. Under the constant flow rate of the buffer solution in the buffer chamber, the radioactivity in the buffer chamber becomes constant in a matter of minutes when the steady state is reached, which can be a measure of the concentration of the free ligand in the sample chamber. Then small amounts of the unlabeled ligand in a small volume are added to the sample chamber, a new chemical equilibrium is attained together with a rapid exchange between isotopes, and the free ligand diffuses at a different rate depending on the concentration in the sample chamber giving a new steady-state level of radioactivity in the effluent. After successive additions of the unlabeled ligand, followed by determinations of the respective steady-state levels of the radioactivity, excess unlabeled ligand in a small volume is added, a maximum value for the radioactivity ($C_n$) in the effluent is reached. This can correspond to that expected when no appreciable fraction of the labeled ligand is bound. That is, this value becomes a measure of the total concentration of the

radioactive ligand in the sample chamber under the conditions where the total unavoidable loss of the radioactive ligand from the sample chamber is sufficiently small and can be neglected. Under such conditions, the concentration of free ligand ($L_i$) in the protein solution at each titration step is estimated from the steady-state value of radioactivity ($C_i$) in the effluent and the known value of total concentration of the ligand ($Lt_i$), because $L_i/Lt_i = C_i/C_n$. Then the concentration of the bound ligand can be calculated. As a result, a complete ligand-binding curve can be obtained through a simple titration experiment on a single protein solution within an hour.

Therefore, once the flow dialysis apparatus necessary is constructed, the method rapidly yields a reliable $Ca^{2+}$-binding curve in about an hour, and the $Ca^{2+}$-binding protein can be characterized extracting the $Ca^{2+}$-binding constants and the number of $Ca^{2+}$-binding sites from computer-aided curve-fitting giving the best-fit curve ***(6–10)***. In this chapter, a protocol for the flow-dialysis method is shown, which illustrates the apparatus that has been used for measurement of $Ca^{2+}$-binding to calmodulin in our laboratory ***(8,11)*** (*see* **Notes 1** and **2**).

## 2. Materials

1. Flow-dialysis cell: The flow-dialysis apparatus used in our laboratory is a Reichard-type flow-dialysis cell ***(5)***, which consists of two blocks, a cylindrical upper chamber (sample chamber) and the lower chamber (buffer chamber). The chambers are separated by a sheet of cellulose membrane that is clamped by the two blocks held with four plastic screws (*see* **Fig. 1**). The Reichard-type cell is made of Teflon and the precise shop drawing is shown in **Fig. 2**. The cross-sectional area of the cylindrical chamber is 1.32 $cm^2$, and a capacity of the lower chamber is 0.66 mL, whereas the upper chamber can accommodate up to 2.5 mL of the sample solution. After assembling the cell, it is immersed in the water bath made of glass, and the water at a constant temperature is circulated through its water jacket (*see* **Fig. 3**). The whole assembly is placed on a magnetic stirrer (*see* **Notes 3** and **4**).
2. Thermostated water bath equipped with a circulator.
3. Magnetic stirrer and Teflon-covered magnetic stirring bars.
4. Peristaltic pump.
5. Fraction collector and plastic tubes (1.1 × 7.4 cm).
6. Liquid scintillation counter.
7. Atomic absorption spectrometer.
8. Dialysis membrane: A square cut, 1.9 × 1.9 cm, or preferably a circular cut with a diameter of 1.9 cm from cellulose dialysis tubing is prepared for use as a dialysis membrane. Select the appropriate membrane considering the rate of dialysis, as well as the pore size to keep proteins and effectors other than the ligand in the sample chamber. Spectra Pore 6 cellulose membrane with molecular cutoff 1000 (Spectrum Industry Inc., Los Angels, CA) is suitable for the purpose of calmodulin ($M_r$ 16,700) and its complex with target peptides ($M_r$ around 3000) ***(8)***.

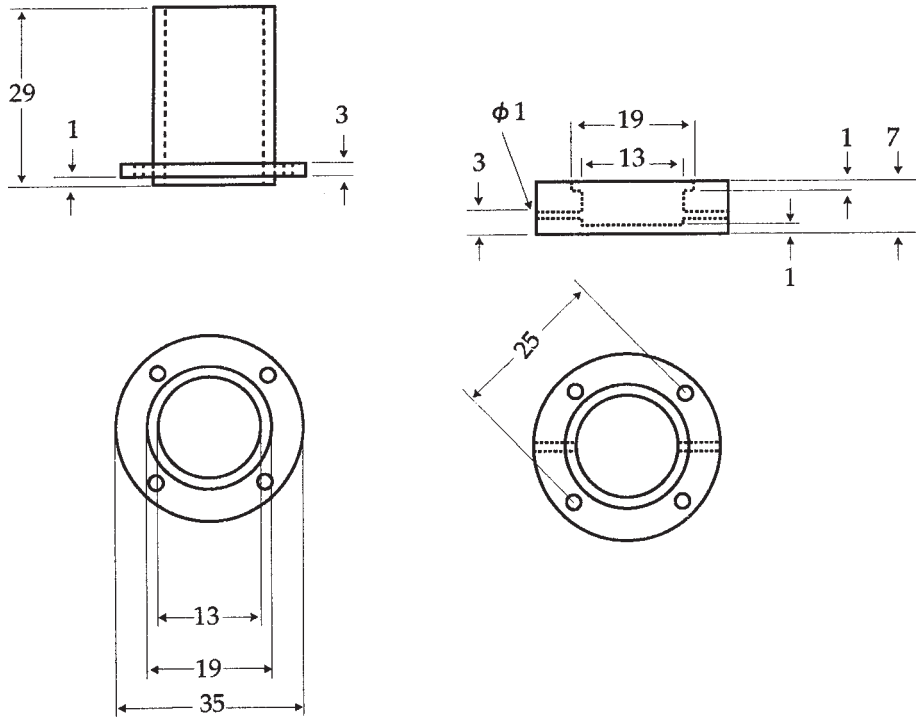


Fig. 2. Shop drawing for the Reichard-type flow-dialysis cell. Left, the upper chamber; right, the lower chamber. Dimensions are shown in millimeters. The flow-dialysis cell is made of Teflon.

The membranes are heated to approx 80°C on the boiling water bath for approx 30 min once in 1 m*M* ethylenediaminetetracetic acid (EDTA), twice in deionized water, then in 1% acetic acid, and twice in deionized water. The washed membranes in deionized water are stored in the refrigerator.

9. Buffer solution: A suitable buffer solution (approx 500 mL) is prepared using deionized distilled water or a Milli-Q water. The buffer solution is preferable to contain 50 m*M* or higher concentration of NaCl or KCl. It should be degassed sufficiently with stirring under reduced pressure using an aspirator immediately before the measurement, which is essential for preventing formation of bubbles in the buffer chamber.
10. Stock solutions of $Ca^{2+}$: The stock solutions of $Ca^{2+}$ used in the $Ca^{2+}$ titration are prepared from $CaCl_2 \cdot 2H_2O$ by weight and stored in plastic bottles (*see* **Note 5**). The $Ca^{2+}$ concentration can be determined by atomic absorption spectrometry.
11. Radioactive $Ca^{2+}$ ($^{45}Ca^{2+}$): The radioactive isotope available for Ca, $^{45}Ca$, is a β-emitting nuclide with a half-life of 164 d and a maximum energy of 0.254 MeV. We purchase $^{45}CaCl_2$ solution with sufficiently high specific radioactivity from

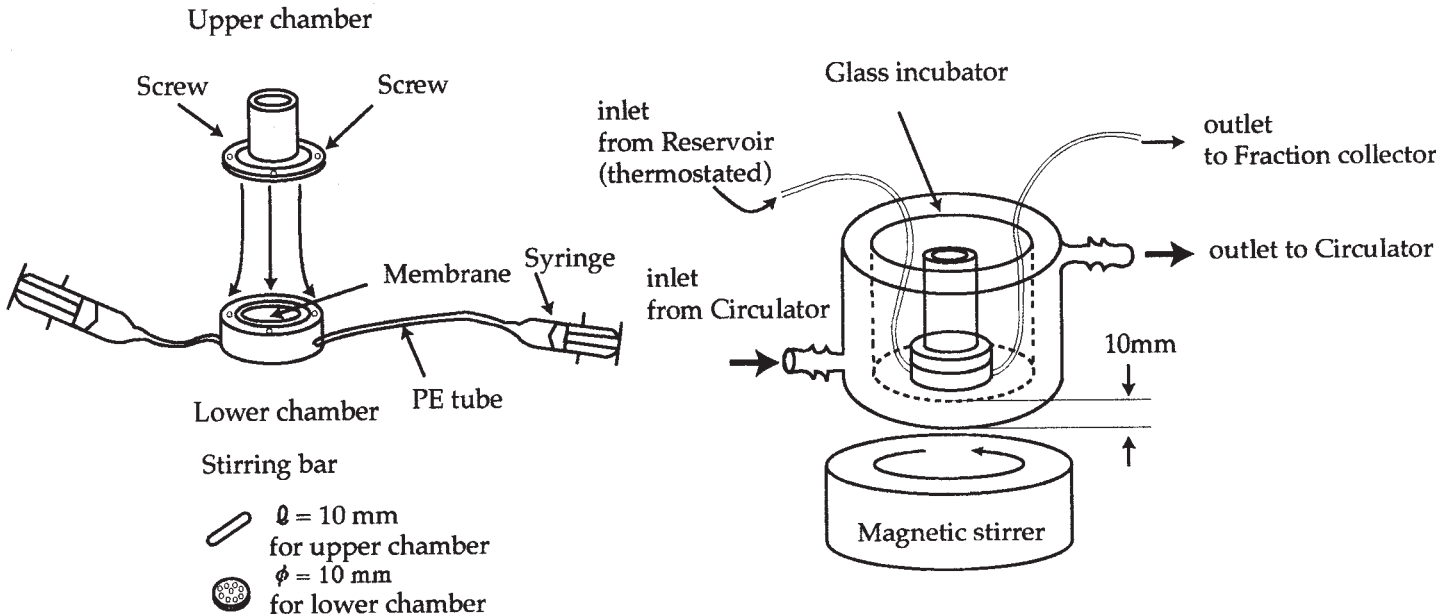


Fig. 3. After assembling the flow-dialysis cell (left), it is equilibrated in the glass incubator (right).

DuPont-NEN and the molar concentration of it is calculated from the specification data shipped with the radionuclide.

12. Protein solution: Contaminating $Ca^{2+}$ in the protein solution should be reduced to less than 0.1 mol/mol of the total $Ca^{2+}$-binding sites in the initial sample solution. Several methods have been reported to reduce the contaminating $Ca^{2+}$ in the neutral pH such as addition of 50–100 m*M* EDTA or ethyleneglycol-*bis* (2-aminoethyl ether)-*N*,*N*,*N*',*N*'-tetraacetic acid (EGTA) (optionally in the presence of 6 *M* urea) followed by passage through a Sephadex G-25 desalting column ***(12)***, and extensive dialysis against or column chromatography with Chelex-100 chelating resin ***(2)***. For stable proteins such as calmodulin, troponin C, and parvalbumin, addition of trichloroacetic acid to 3% precipitates $Ca^{2+}$-free proteins, which can be renatured by neutralization ***(6,13)***. Precipitates were collected by centrifugation, homogenized with addition of small volume of deionized water, and dissolved with addition of small volume of 2 *M* Tris base, which is followed by passage through a Sephadex G-25 column. The concentration of contaminating $Ca^{2+}$ in the protein solution is determined by atomic absorption spectrometry, and is taken into account in the calculation of the $Ca^{2+}$-binding data.

Protein concentrations must be determined with a suitable and reliable method, such as quantitative amino acid analysis or UV spectro-photometric determination.

## 3. Methods

The following procedures, except for the final calculation step, as well as the dilution of $^{45}Ca^{2+}$ aforementioned, are to be carried out in a room equipped with facilities necessary for handling of radioisotopes.

### *3.1. Assembly of the Flow-Dialysis Cell*

When everything is ready to begin, the flow-dialysis cell, which has been stored as disassembled parts, is assembled with mounting a dialysis membrane (*see* **Fig. 3**) (*see* **Note 6**).

1. Immerse the degassed buffer solution in the 500 mL Erlenmeyer flask in the water bath and equilibrate at 25°C.
2. Fill a disposable plastic syringe (5 mL) with the buffer solution and, after removal of air bubbles, connect it to the PE tube (inlet) of the buffer chamber of the dialysis cell.
3. Fill the buffer chamber with the buffer solution sent from the syringe through the PE tube.
4. Fill the other PE tube (outlet) with the buffer solution from the chamber by sucking with another syringe connected to the other end of PE tube.
5. Put a magnetic stirring bar (discoid, 1 cm in diameter) into the chamber, and fill the chamber with the buffer solution.

6. Take out a washed dialysis membrane, trim away the corners of the square cut to fit within a circle (1.9 cm in diameter), and rinse it with deionized water (handle the membrane with clean tweezers).
7. Cover the solvent surface in the buffer chamber with the rinsed dialysis membrane (handle with clean tweezers).
8. Remove the air bubbles, if any, in the buffer chamber using a 200-μL pipeteman, and fit the membrane on the Teflon surface on which the bottom of the upper chamber sits to clamp the membrane.
9. Set the upper chamber gently adjusting carefully the positions of the screw holes.
10. Hold the two blocks together with four plastic screws driven evenly.
11. Remove the buffer solution overflowed into the upper chamber during the assembling with use of a 200-μL pipeteman.

### 3.2. Flow Dialysis

1. Immerse the assembled dialysis cell in a glass incubator, which is placed on a magnetic stirrer and has been equilibrated at 25°C by circulating the thermostated water.
2. Remove syringes at the tips of PE tubings after pinching them with the hemostat, and connect one tip (inlet) to the end of a tube filled with the buffer leading from the buffer reservoir through the peristaltic pump and the other tip (outlet) to the drain.
3. Release the hemostats, start the magnetic stirrer and the peristaltic pump at a constant flow rate of 1–3 mL/min (3 mL/min is recommended), and watch the buffer chamber to confirm that no air bubbles are formed.
4. Switch off the magnetic stirrer and put into the upper chamber a magnetic stirring bar (rod, 0.1 cm in diameter and 1 cm in length).
5. Add 1.5 mL of the protein solution to the upper chamber, and resume stirring. About 5–10 times the anticipated $K_d$ value is recommended for the initial concentration of the $Ca^{2+}$-binding protein ***(5)***. In the case of calmodulin, the apparent $K_d$ is around 5 μ*M* and an initial concentration of 20–100 μ*M* is suitable for the measurement.
6. Equilibrate the solutions for about 5 min while stirring both chambers and continuously flushing of the buffer chamber.
7. Take out 0.5 mL of the protein solution from the upper chamber, transfer it into a microtube and store on ice for an exact determination of the protein and contaminating $Ca^{2+}$ concentrations later.
8. Connect the effluent tip to the fraction collector, and start to collect the effluent (1 mL/plastic tube) continuously through the end of titration.
9. At tube number 7, add 5 μL of $^{45}Ca^{2+}$ with sufficient specific radioactivity (1–5 × $10^7$ cpm, and the amount of $Ca^{2+}$ equal to about 1/10 of the total number of the $Ca^{2+}$-binding site) to the upper chamber.
10. At tube number 13, add 5 μL of the unlabeled $Ca^{2+}$ (the amount of $Ca^{2+}$ equal to about 1/10 of the total number of the $Ca^{2+}$-binding site).

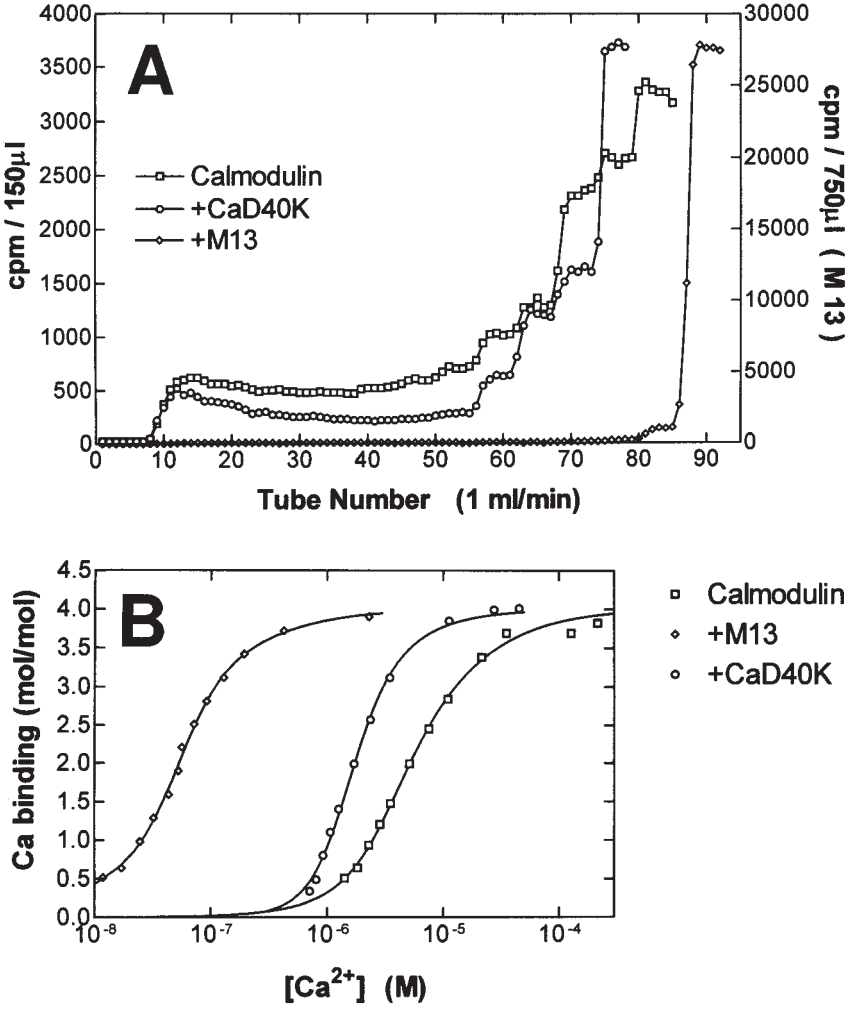


Fig. 4. Examples of the $Ca^{2+}$ titration curves for the measurements of $Ca^{2+}$-binding to calmodulin **(A)**, and the resulting $Ca^{2+}$-binding curves **(B)**. Flow dialysis was carried out on smples of 16.5 µ*M* calmodulin (□), 0.1 *M* NaCl, and 0.02 *M* MOPS-NaOH (pH 7.0) at 25°C at a flow rate of 1 mL/min. Measurements were also carried out in the presence of target peptide: the 40-kDa fragment of caldesmon (CaD40K, ○; 28.5 µ*M*), and M13 peptide from myosin light chain kinase (◇; 43.9 µ*M*). In each titration, $^{45}Ca^{2+}$ ($6.67 \times 10^6$ cpm for □, ○, and $9.73 \times 10^6$ cpm for ◇) was added at tube number 7, then at every six collected tubes, unlabeled $Ca^{2+}$ was added successively. During the titration, 1.5–6.6% of $^{45}Ca^{2+}$ in the sample chamber was lost. The interaction of calmodulin with CaD40K increases the cooperativity of $Ca^{2+}$-binding, whereas the interaction with the M13 peptide markedly increases the affinity for $Ca^{2+}$ **(B)**, which is differently reflected in the titration curves **(A)**

11. Similarly, add 5 to 10 μL of the unlabeled $Ca^{2+}$ successively at every six tubes collected. The recommended total amount of $Ca^{2+}$ to be added at each step of the successive six steps is 2/10 of the total number of the $Ca^{2+}$-binding site, which covers to 1.4 times the total $Ca^{2+}$-binding sites.
12. Finally, add 10 μL of 1 *M* $CaCl_2$ to chase practically all of the bound $^{45}Ca^{2+}$, and collect six more tubes, then switch off the pump and the fraction collector. Connect the effluent tip to the bottle for the waste $^{45}Ca^{2+}$.
13. Take out a constant volume of the effluent in each collected tube and quantify $^{45}Ca^{2+}$ with the liquid scintillation counter. Make a titration curve to confirm that steady-state has been reached at each titration step as shown in **Fig. 4A**.

### 3.3. Disassembly of the Flow-Dialysis Cell

1. Start the peristaltic pump to wash out the radioactive solvent in the lower chamber with the fresh solvent. Collect the radioactive effluent into the bottle for waste $^{45}Ca^{2+}$. Then switch off the pump.
2. Take out the radioactive sample solution in the upper chamber and transfer into the bottle for waste $^{45}Ca^{2+}$ with use of a Pasteur pipet.
3. Add a small amount of the detergent solution into the sample chamber, rinse with it the inner surface and transfer the resulting radioactive solution into the bottle for waste $^{45}Ca^{2+}$. Repeat at least three times to remove the radioactivity.
4. Take out the dialysis cell from the incubator, place it on the bench, and discharge the solvent in the lower chamber into the bottle for waste $^{45}Ca^{2+}$.
5. Disassemble the apparatus carefully with releasing screws. With tweezers, put the upper and lower chambers and stirring bars into the detergent solution, and the dialysis membrane into the can for the radioactive waste.
6. Wash the disassembled parts thoroughly with detergent solution, rinse with the distilled water.

### 3.4. Calculation to Make a $Ca^{2+}$-Binding Curve

1. Average the steady-state values of the radioactivity usually obtained in two or three tubes just prior to the addition of next $Ca^{2+}$ and subtract the averaged baseline value obtained before the initial addition of $^{45}Ca^{2+}$. Then, correct for the dilution at each step of titration to yield the net average value of radioactivity, $C_1$, $C_2$,….., $C_n$ at each step of titration.
2. Considering the dilution factor again, calculate the total concentration of $Ca^{2+}$; $Ca_1$, $Ca_2$,……., $Ca_n$ at each step from the amounts of added $Ca^{2+}$ and initial concentration of $Ca^{2+}$ in the protein solution determined by atomic absorption spectrometry.
3. Calculate the concentration of free $Ca^{2+}$ in the upper chamber at each step from $[Ca^{2+}]^{free} = Ca_iC_i/C_n$.
4. Calculate concentration of bound $Ca^{2+}$ from the difference between concentrations of total $Ca^{2+}$ and free $Ca^{2+}$ at each step, which gives a molar ratio of bound $Ca^{2+}$ to the $Ca^{2+}$-binding protein ($Ca^{2+}$-binding number) considering the dilution factor in the calculation of the concentration of $Ca^{2+}$-binding protein.

5. Plot the calculated $Ca^{2+}$-binding number against pCa = –Log$[Ca^{2+}]^{free}$ as shown in **Fig. 4B**.

The resulting $Ca^{2+}$-binding curve (*see* **Fig. 4B**) can be analyzed by curve fitting based on several different models of $Ca^{2+}$-binding equilibrium ***(2,6–12)***, details of which are described in Chapter 3 by Haiech.

## 4. Notes

1. The method is based on the following assumptions: (1) The rate at which the ligand leaves the sample chamber is proportional to the concentration of free ligand. (2) $Ca^{2+}$-binding equilibrium in the sample chamber is attained rapidly compared with the response time of the apparatus. (3) $Ca^{2+}$ bound to the protein in the sample chamber can be exchanged rapidly with the free $Ca^{2+}$ compared with the response time ***(4,5)***. Under these conditions, the rate of change in the number of $Ca^{2+}$ in the lower chamber can be given by $dN/dt = L_iD–Nv/V$, because the rate of diffusion of $Ca^{2+}$ across the membrane into the lower chamber is given by the product of its concentration ($L_i$) and a constant determined by the properties of membrane ($D$), and the rate of exit of $Ca^{2+}$ (initially N molecules present) from the lower chamber is determined by its volume ($V$) and the flow velocity ($v$) of the buffer solution. Therefore, at the steady state, the concentration of $Ca^{2+}$ in the lower chamber, which is given by $N/V = L_iD/v$, is proportional to the concentration of the free $Ca^{2+}$ in the upper chamber.
2. Practically, the steady state in the lower chamber is attained when the volume $vt$ flowing through the lower chamber is four times the volume of the lower chamber ***(4,5)***, and the required time $t = 4V/v$ is termed a response time of the cell. The response time can be determined experimentally by monitoring the $^{45}Ca^{2+}$ in the effluent (*see* **Fig. 4A**). In our dialysis cell, the volume of the lower chamber is 0.66 mL and when the lower chamber is flushed with a flow rate of 1 mL/min, the response time is 4 × 0.66 = 2.64 min, that is, a new steady-state is attained in 2.64 min after the addition of the ligand. Intervals of adding $Ca^{2+}$ in the titration can be estimated from the response time, which is determined by the flow rate and the volume of the lower chamber.
3. As indicated by the basic equation shown in **Note 1**, $^{45}Ca^{2+}$ detectable in the effluent may increase with decrease in the flow rate, and seems favorable for the measurement. We have, however, another basic assumption: (4) Total amounts of ligand diffused out into the buffer chamber during the whole titration process are small and can be neglected ***(4,5)***. Considering this basic assumption, gaining high signal by decreasing the flow rate is incorrect because $^{45}Ca^{2+}$ in the upper chamber may be lost too much during the whole titration process. $^{45}Ca^{2+}$ with higher specific activity should be used for this purpose. Similarly unnecessary repetitive titration at highly saturating concentrations of $Ca^{2+}$ must be avoided. Because with a given apparatus $^{45}Ca^{2+}$ detected in the effluent is determined by the flow rate, one should confirm it experimentally and set up the experimental flow rate and repetitive numbers of titration considering the amount of unavoidable loss. In our

apparatus, 0.42 and 0.10% of $^{45}Ca^{2+}$ in the upper chamber is detected in the effluent at a flow rate of 1 mL/min and 3.3 mL/min, respectively, which makes up the loss.

4. When the loss of more than 10% of the initial $^{45}Ca^{2+}$ cannot be prevented because of other reasons, experimental result must be corrected for the loss, which can be estimated by summing up the radioactivity in the effluent. Details for the correction are described by Stemmer and Klee ***(10)***.
5. Plastic containers are recommended to keep solutions. Plasticware may be soaked in 1 *M* HCl for several hours to eliminate contaminating $Ca^{2+}$ and rinsed thoroughly with deionized water.
6. A more efficient model of Reichard-type flow-dialysis cell has been constructed by Porumb ***(5,9)***. Another advanced model of the flow-dialysis cell has been constructed by Feldmann ***(6,14)***. In the Feldmann-type cell, the volume of the lower chamber is minimized by engraving a spiral groove on the surface of a solid dome and the loss of $^{45}Ca^{2+}$ can be minimized. Unfortunately, it was a little difficult for our shop to construct it.

## References


1. Potter, J. D. and Gergely, J. (1975) The calcium and magnesium binding sites on troponin and their role in the regulation of myofibrillar adenosine triphosphatase. *J. Biol. Chem.* **250,** 4628–4633.
2. Crouch, T. H. and Klee, C. B. (1980) Positive cooperative binding of calcium to bovine brain calmodulin. *Biochemistry* **19,** 3692–3698.
3. Teraoka, H. and Nierhaus, K. H. (1979) Measurement of the binding of antibiotics to ribosomal particles by means of equilibrium dialysis. *Methods Enzymol.* **59,** 862–866.
4. Colowick, S. P. and Womack, F. C. (1969) Binding of diffusible molecules by macromolecules: rapid measurement by rate of dialysis. *J. Biol. Chem.* **244,** 774–777.
5. Womack, F. C. and Colowick, S. P. (1973) Rapid measurement of binding of ligands by rate of dialysis. *Methods Enzymol.* **27,** 464–471.
6. Haiech, J., Klee, C. B., and Demaille, J. G. (1981) Effects of cations on affinity of calmodulin for calcium: ordered binding of calcium ions allows the specific activation of calmodulin-stimulated enzymes. *Biochemistry* **20,** 3890–3897.
7. Minowa, O. and Yagi, K. (1984) Calcium binding to tryptic fragments of calmodulin. *J. Biochem.* **56,** 1175–1182.
8. Yazawa, M., Ikura, M., Hikichi, K., Luan, Y., and Yagi, K. (1987) Communication between two globular domains of calmodulin in the presence of mastoparan or caldesmon fragment. *J. Biol. Chem.* **262,** 10,951–10,954.
9. Porumb, T. (1994) Determination of calcium-binding constants by flow dialysis. *Anal. Biochem.* **220,** 227–237.
10. Stemmer, P. M. and Klee, C. (1994) Dual calcium ion regulation of calcineurin by calmodulin and calcineurin B. *Biochemistry* **33,** 6859–6866.
11. Yazawa, M., Vorherr, T., James, P., Carafoli, E., and Yagi, K. (1992) Binding of calcium by calmodulin: influence of the calmodulin binding domain of the plasma membrane calcium pump. *Biochemistry* **31,** 3172–3176.

12. Starovasnik, M. A., Davis, T. N., and Klevit, R. E. (1993) Similarities and differences between yeast and vertebrate calmodulin: an examination of the calcium binding and structural properties of calmodulin from the yeast *Saccharomyces cerevisiae*. *Biochemistry* **32,** 3261–3270.
13. Yazawa, M., Sakuma, M., and Yagi, K. (1980) Calmodulins from muscles of marine invertebrates, scallop and sea anemone. *J. Biochem.* **87,** 1313–1320.
14. Feldmann, K. (1978) New devices for flow dialysis and ultrafiltration for the study of protein-ligand interactions. *Anal. Biochem.* **38,** 225–235.

# 2

# Calcium Binding to Proteins Studied via Competition with Chromophoric Chelators

**Sara Linse**

## 1. Introduction

Optical spectroscopic techniques can be used to measure $Ca^{2+}$-binding constants when the $Ca^{2+}$-bound and free forms of the protein display a difference in, for example, the UV absorbance, CD or fluorescence spectrum, or fluorescence polarization. One may then start with the $Ca^{2+}$-free form, titrate in $Ca^{2+}$ stepwise, measure a spectrum or intensity at each step, and obtain the binding constants from computer fitting to the data. The best accuracy is achieved when the protein concentration is roughly the same as the dissociation constant (the inverse of the binding constant) such that there are significant populations of both bound and free forms at several titration points. This limits the useful range of such direct measurements to binding constants below $10^6$ $M^{-1}$ ($K_D >$ 1 $\mu M$), because of the practical difficulty of making buffers with less than 0.5–1 $\mu M$ free $Ca^{2+}$. For $Ca^{2+}$-binding proteins with affinities of $10^6$ $M^{-1}$ and up, one has to rely on indirect measurements. One popular such approach uses around 1 m$M$ ethylenediaminetetracetic acid (EDTA) or ethylene glycol-*bis* *N*,*N*,*N*',*N*'-tetraacetic acid (EGTA), and a much smaller amount of protein so that the free-$Ca^{2+}$ concentration is essentially controlled by the $Ca^{2+}$-buffering capacity of EDTA or EGTA. A potential risk with such approaches is binding of EDTA or EGTA to the protein with consequences for its $Ca^{2+}$ affinity. Another type of indirect approach outlined in this chapter involves the use of a chelator whose absorbance or fluorescence is $Ca^{2+}$ dependent ***(1–3)***. A mixture of equal (10–50 $\mu M$) amounts of chelator and protein is titrated with $Ca^{2+}$ and the binding to the chelator is monitored spectroscopically. The $Ca^{2+}$-binding constants of the protein are extracted by fitting to the absorbance or fluores-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

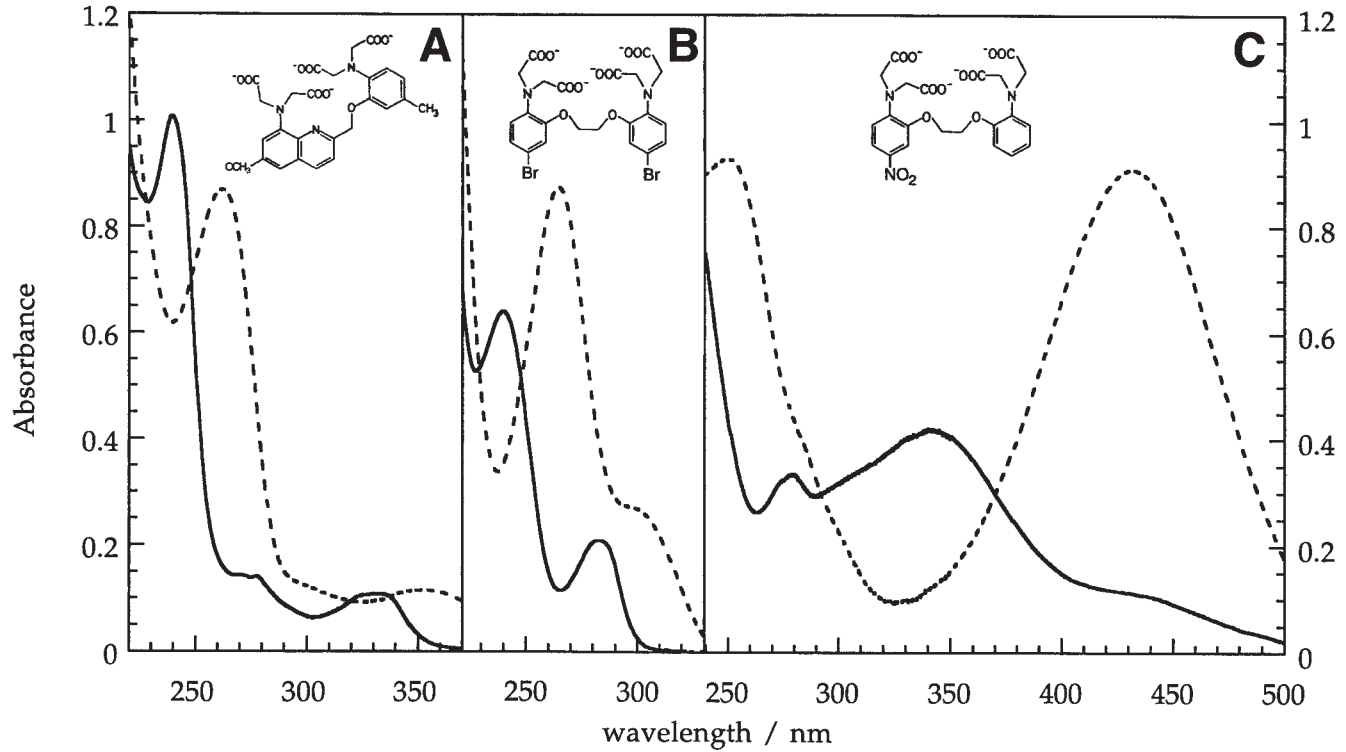


Fig. 1. Molecular structures and absorbance spectra of **(A)** quin-2; **(B)** 5,5'$Br_2$-BAPTA; and **(C)** 5N-BAPTA. (- - -) calcium free and (—) calcium bound forms.

**Table 1**
**Molecular Structures, Spectra and Properties of 3 Chelators**

| Chelator | $\lambda_{max}$/nm | ε/*M*/cm | *KD*/M[a] low salt[b] | *KD*/*M* 0.15 *M* KCl | *KD*/*M* 0.15 *M* NaCl | *Mw* | ref. |
|---|---|---|---|---|---|---|---|
| quin-2 | 239.5 | $4.2 \cdot 10^4$ | $5.2 \cdot 10^{-9}$ | $1.2 \cdot 10^{-7}$ | | 694[c] | ***1,2*** |
| 5,5'$Br_2$-BAPTA | 239.5 | $1.4 \cdot 10^4$ | $1.0 \cdot 10^{-7}$ | $2.3 \cdot 10^{-6}$ | $1.4 \cdot 10^{-6}$ | 787[c] | ***1–4*** |
| 5N-BAPTA | 340 | $6.0 \cdot 10^3$ | $1.7 \cdot 10^{-6}$ | | $2.7 \cdot 10^{-5}$ | 521[d] | ***1,4*** |

[a]All KD's are in 2 m*M* Tris-HCl at pH 7.5.
[b]No salt added beyond the HCl needed to set the pH.
[c]Tetra potassium salt.
[d]Free acid.
Quin-2 can be obtained from Fluka, Buchs, Switzerland, and 5,5'$Br_2$-BAPTA and 5N-BAPTA from Molecular Probes, Eugene, OR.

cence as a function of total $Ca^{2+}$ concentration. This method gives very high precision in the deduced constants, but the accuracy is never better than the accuracy in the $Ca^{2+}$ affinity for the chelator. Although much lower concentrations of chelator are used, this method is also potentially hampered by interactions between chelator and protein. Another source of errors are electrostatic screening effects from highly charged proteins that perturb the calcium affinity for the chelator from its value in a protein-free solution.

## 2. Materials

1. UV absorbance or fluorescence spectrometer.
2. Quartz cuvets.
3. Chromophoric calcium chelator. An ideal chelator is one with a calcium affinity close to that of the protein to be studied. This will ensure that the calcium ions are roughly evenly distributed between the chelator and protein leading to high precision in the binding constants for the protein. The molecular structures, spectra and properties of three useful chelators are summarized in **Fig. 1** and **Table 1**.
4. $Ca^{2+}$-free buffer (*see* **Note 1**). To get the buffer $Ca^{2+}$ free, prepare in double-distilled water ($ddH_2O$) in a plastic container and put a dialysis tube filled with Chelex-100 resin (Bio-Rad) in the container before adjusting the pH (*see* **Note 2**). Before use, the dialysis tube has to be boiled four times in $ddH_2O$ and the chelex has to be neutralized and washed with $ddH_2O$. Let the buffer rest for a few days before use to reduce free $Ca^{2+}$.
5. 3 m*M* $CaCl_2$. Weigh as accurately as you can 44.106 mg $CaCl_2 \cdot 2H_2O$ (*see* **Note 3**). Note the exact weight and calculate the $Ca^{2+}$ concentration from that value. Dissolve the $Ca^{2+}$-free buffer in a 100-mL volumetric flask. Adjust the pH, if necessary, and fill up the flask. Aliquot into a large number of Eppendorf tubes and freeze the tubes. For each titration, use one tube and then dispose.
6. 1 *M* $CaCl_2$. Dissolve 14.72 g $CaCl_2 \cdot 2H_2O$ in 100 mL $ddH_2O$ and adjust pH to 7.5.

7. 0.1 *M* EDTA. Dissolve 37.22 g EDTA in 100 mL $ddH_2O$. Add concentrated NaOH to get the EDTA into solution and adjust the pH to 7.5.
8. 5 m*M* EDTA. Dilute 25 mL 0.1 *M* EDTA with 475 mL $ddH_2O$ in a squeeze bottle.

## 3. Method

### 3.1. Experimental Procedure

1. A $Ca^{2+}$-free solution of 25–30 μ*M* chelator is prepared in the $Ca^{2+}$-free buffer. The exact chelator concentration $C_Q$ is determined by withdrawing 2.5 mL, adding 5 μL 1 *M* $CaCl_2$ and recording the absorbance at $\lambda_{max}$ (*see* **Table 1**). The chelator concentration is calculated as $CQ = A_{\lambda max}/\varepsilon$. The value of ε at $\lambda_{max}$ is found in **Table 1**.
2. Rinse the cuvet once with $ddH_2O$. Fill with 5 m*M* EDTA and let sit for 1 min. Rinse several times with $ddH_2O$ and finally with ethanol and dry the cuvet with nitrogen gas.
3. Record the absorbance at 263 nm (*see* **Note 4**) $A_{263}$ for 2.5 mL of the chelator solution (-> $A_1$). Add 5 μL 0.1 *M* EDTA and record $A_{263}$ (-> $A_2$). Add 5 μL 1 *M* $CaCl_2$ and record $A_{263}$ (-> $A_3$). The calcium concentration in the chelator solution $Ca_Q$ can be estimated as

$$Ca_Q = C_Q \cdot (A_2 - A_1) / (A_2 - A_3) \quad (1)$$

   Ideally, this value is below 1 μ*M* (*see* **Note 5**).
4. Rinse the cuvet once with $ddH_2O$. Fill with 5 m*M* EDTA and let sit for 1 min. Rinse several times with $ddH_2O$ and, finally, with ethanol, and dry the cuvet with nitrogen gas.
5. Dissolve lyophilized $Ca^{2+}$-depleted protein (*see* **Note 6**) in the ($Ca^{2+}$- and EDTA-free) chelator solution to obtain a protein concentration of 25–30 μ*M*. This is the titrand, i.e., the solution that will be titrated with calcium.
6. Record $A_{263}$ (*see* **Note 4**) for the titrand.
7. Add a $Ca^{2+}$ aliquot (*see* **Note 7**) to the titrand and mix. Record $A_{263}$ (*see* **Note 8**).
8. **Step 7** is repeated until no significant change has occurred in $A_{263}$ over the last five points, beyond what would be caused by dilution (*see* **Note 9**).

### 3.2. Computer Fitting

The chelator method can be used to determine *macroscopic* $Ca^{2+}$-binding constants of a protein. Because the measured quantity contains no information about the distribution of calcium among separate sites in the protein, *microscopic* binding constants cannot be determined. The *macroscopic* binding constants $K_1$, $K_2$–$K_N$ are defined as follows:

$$\begin{aligned} &P + Ca^{2+} \rightleftharpoons PCa && K_1 = [PCa]/([P][Ca^{2+}]) \\ &PCa + Ca^{2+} \rightleftharpoons PCa_2 && K_2 = [PCa_2]/([PCa][Ca^{2+}]) \\ &\ldots \\ &PCa_{N-1} + Ca^{2+} \rightleftharpoons PCa_N && K_N = [PCa_N]/([PCa_{N-1}][Ca^{2+}]) \end{aligned} \quad (2)$$

$K_1$–$K_N$ (where $N$ is the number of sites that are strong enough to compete with the chelator) are obtained by nonlinear least squares fitting to the absorbance as a function of total calcium concentration. An analysis based on concentration (not activities) can be performed as follows (*see* **Note 10**).

The total $Ca^{2+}$-concentration at each titration point $i$ ($CATOT_i$), is calculated from the initial (*see* **Note 11**) and added $Ca^{2+}$. A nominal value for the protein concentration at each titration point ($CP_i$) is calculated from the initial protein concentration based on the weight of the lyophilized protein. $CATOT_i$ and $CP_i$ are adjusted for the dilution imposed by the calcium additions, as is $CQ_i$, the chelator concentration at titration point $i$. Fixed parameters in the fit are KDQ, $CQ_i$, $CP_i$, and $CATOT_i$. KDQ is the $Ca^{2+}$-dissociation constant of the chelator. Variable parameters in the fit are $K_1$–$K_N$, AMAX, AMIN, and $F$. AMAX and AMIN are the absorbances that the initial (nondiluted) solution would have had if it was completely $Ca^{2+}$-free or contained saturating amounts of $Ca^{2+}$, respectively. $F$ is a correction factor that accounts for the fact that the protein concentration obtained by weight can be off by 10–20% because of residual water in lyophilized protein and because of errors in weight caused by the small (0.7–1.5 mg) quantities used (*see* **Note 12**).

For each set of values of the variable parameters, the Newton-Raphson method is used to solve the free $Ca^{2+}$ concentration, $Y$, at each titration point, $i$, from the following equation:

$$Y = CATOT_i - \frac{CQ_i \cdot Y}{Y + KDQ} - \frac{F \cdot CP_i \sum_{k=1}^{N} \left(kY_k \cdot \prod_{j=1}^{k} K_j\right)}{1 + \sum_{k=1}^{N} \left(Y_k \cdot \prod_{j=1}^{k} K_j\right)} \tag{3}$$

which states that the free $Ca^{2+}$ equals the total $Ca^{2+}$ subtracted by the chelator-bound $Ca^{2+}$ and the protein-bound $Ca^{2+}$. The absorbance at point $i$ is calculated as

$$Acalculated,i = \left[AMAX - (AMAX - AMIN) \cdot \frac{Y}{Y + KDQ}\right] \cdot \frac{CQi}{Q_1} \tag{4}$$

where $CQ_1$ is the initial chelator concentration. Thus the changes in absorbance are assumed to arise from the chelator only. The sum of the squares of residuals (or error square sum) $\chi^2$, is obtained by summing over all points in the titration

$$\chi^2 = \sum (A_{calculated,i} - A_{measured,i})^2 \tag{5}$$

The variable parameters are iterated in a separate procedure until an optimal fit (minimum $\chi^2$) is found. Start with initial guesses at both sides of the parameter values of best fit, to make sure that the same result is obtained. To estimate the errors in the parameter values, one may fix one parameter, for example $K_1$, and

iterate the other parameters to obtain an optimal fit. Then fix $K_1$ at a new value and fit again. Repeat until you have found the values of $K_1$ that lead to a doubling of $\chi^2$. In general, AMAX, AMIN, and $F$ are better determined than the binding constants (*see* **Note 13**). If the protein binds calcium with positive cooperativity (*see* **Note 14**), the product of the binding constants is better determined than the individual constants.

### 3.3. Stoichiometry of Calcium Binding

The chelator method can be used to measure the stoichiometry of calcium binding. For such applications, extra care has to be taken to measure the protein concentration of the titrand and its initial and final calcium concentration.

1. Dissolve the protein in 3 mL chelator solution to approx 30 μ*M*.
2. Withdraw 200 μL. Freeze dry for acid hydrolysis.
3. Use 2.5 mL as titrand.
4. Save the rest for atomic absorption spectroscopy for initial calcium concentration analysis.
5. Record $A_{263}$ for the titrand.
6. Add a calcium aliquot to the titrand and mix. Record $A_{263}$.
7. Repeat **step 6** until no significant $A_{263}$ change has been observed over the last five points.
8. Withdraw an aliquot of the titrated titrand for atomic absorption spectroscopy for calcium analysis.
9. In the computer fitting, set the initial protein concentration to the value obtained from the amino acid analysis, and use a fixed factor $F$ = 1.0. The number of macroscopic binding constants needed to obtain an optimal fit will be the same as the number of sites with affinities of similar value as the chelator. The initial calcium concentration used in the fit is obtained from the analysis at **step 4**. Check that the total calcium concentration at the last titration point is equal to the value obtained from the analysis at **step 8**.

### 3.4. Examples of Titration Data

Examples of experimental data and fitted curves are shown in **Fig. 2**. In the absence of calcium binding to the protein, the absorbance will decrease linearly until the total calcium concentration equals the chelator concentration. A linear decrease will be seen also when the protein has a site with the same $Ca^{2+}$ affinity as the chelator, but more calcium will be needed to saturate the chelator. If the protein binds calcium weaker or stronger than the chelator, the binding curve will be no longer be a straight line, but will bend in a different direction depending on whether the affinity for the protein is higher or lower than for the chelator (*see* **Fig. 2A**). Examples of experimental data for proteins with one, two, or three high-affinity calcium-binding sites are shown in **Fig. 2B**. When the protein binds calcium at more than one site in a sequential manner

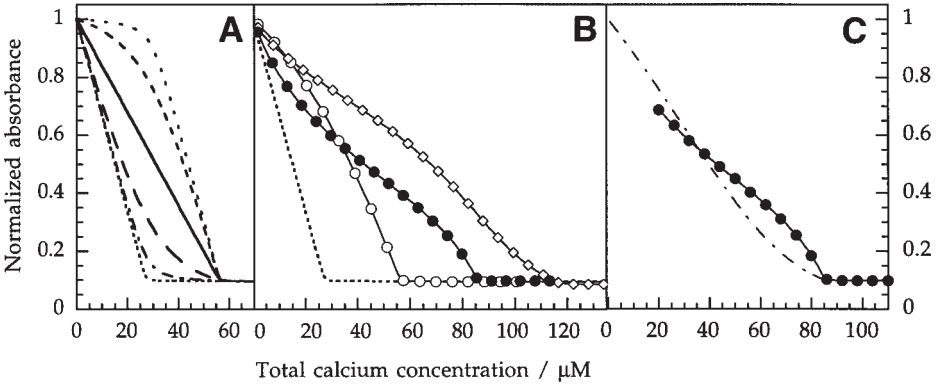


Fig. 2. The absorbance at 263 nm as a function of total calcium concentration for a mixture of 27.5 μ*M* quin-2 and 30 μ*M* protein. **(A)** Simulated curves for three proteins, each with one calcium-binding site with the same (—), 100-fold higher (short dashes), 10-fold higher (dashes), 10-fold lower (long dashes), or 100-fold lower (dash dotted) affinity than quin-2, plus one curve for chelator in the absence of protein (closely spaced short dashes). **(B)** Experimental data (symbols) and fitted curves (solid lines) for three proteins with different stoichiometries of calcium binding: (o) α-lactalbumin, $lgK_1 = 8.7$, (•) calbindin $D_{9k}$ (recombinant bovine minor *A* with a P43M substitution), $lgK_1 = 7.75$. $lgK_2 = 8.59$ ***(5)***,(·) calerythrin $lgK_1 = 8.08$. $lgK_2 = 9.10$, $lgK_3 = 7.57$ ***(6)***. Simulated curve for chelator alone (dashed line). **(C)** (•) experimental data for calbindin $D_{9k}$ contaminated with 20 μ*M* $Ca^{2+}$ (the initial nondefined part of the fitted line is omitted). Simulated curve (dash dotted) for a protein with $lgK = 7.5$, contaminated with 30 μ*M* EDTA.

the titration curve may be S-shaped. Positive cooperativity (*see* **Note 14**) of $Ca^{2+}$-binding is also manifested as an S-shape in the titration curve as observed for calbindin $D_{9k}$ (**Fig. 2B**,**C**), but the curvature is opposite to that of sequential binding. Calerythrin (**Fig. 2B**) first binds calcium to two sites with positive cooperativity, and then to a third weaker site, which is seen as two interlocked and opppositely bent Ss. As illustrated in **Fig. 2C**, a contamination with EDTA may be observed as an initial strong phase that may not fit with the protein concentration. A contamination with calcium leads to loss of data in the beginning of the curve (*see* **Fig. 2C** and **Note 15**).

## 4. Notes

1. Buffers and pH. Examples of useful buffers are 2 m*M* Tris-HCl, pH 7.5, for low ionic-strength measurements or 2–10 m*M* Tris-HCl, pH 7.5, with added salt for higher ionic-strength conditions. The choice of pH depends on the $pK_a$ values of titrable groups on the chelator, as well as on the protein. It is best to use a pH significantly far from any $pK_a$-values, so that the binding constants are not sensi-

tive to small alterations in pH. For 2-[[2-[*bis*(carboxymethyl)amino]-5-methylphenoxy]methyl]-6-methoxy-8-[*bis*-(carboxymethyl)-amino]quinoline (quin-2), the highest $pK_a$ value is 6.36, and for 5,5'-dibromo-1,2-*bis*(O-amino-phenoxy)-ethane-*N*,*N*,*N*,'*N*'-tetraacetic acid (5,5'-Br2BAPTA) it is 5.6 ***(1)***.

2. The chelex tube may shift the pH of the buffer. The buffer may also slowly adjust after the pH has been changed by several units. It is often wise to avoid bringing the pH all the way to the goal. Instead, stop 0.5 pH units above or below (at the side from where you start). Do the final adjustment after a day or two. Your HCl or NaOH stock may contain some calcium so it is often best to wait up to a week before using the buffer.
3. Beware that calcium chloride is hygroscopic.
4. Choice of wavelength. The method is of course not limited to measurements at 263 nm. The ideal wavelength is one at which the calcium induced absorbance change for the chelator is as large as possible while the absorbance for the protein is calcium-independent. For the chelators quin-2 (*see* **Fig. 1A**) and 5,5'$Br_2$-BAPTA (*see* **Fig. 1B**), the absorbance at 263 nm decreases as a consequence of $Ca^{2+}$-binding. **Equation 4**, of course, pertains equally well to a case with increasing absorbance, e.g., another chelator and/or another wavelength. For 5N-BAPTA (*see* **Fig. 1C**) it is convenient to use 430 nm where the protein has no absorbance.
5. If the calcium concentration is not below 1 μ*M*, the buffer may need to rest for a few days to reduce free calcium, or maybe the chelator solution was contaminated with calcium by accident. Sometimes it seems as if solutions get calcium contaminated if you just look at them. Common sources of contamination are pipet tips, pH electrode, and glassware.
6. If the protein cannot be lyophilized, it may be added from a stock solution and the concentration of the chelator solution adjusted for the dilution. The use of a stock solution, however, necessitates the use of one extra container, e.g., Eppendorf tube, which may lead to calcium contamination. The safest procedure is to weigh out lyophilized protein in a cuvet that has been washed with 5 m*M* EDTA, multiple times with $ddH_2O$, finally, with ethanol, and then dried. One procedure to decalcify a high affinity $Ca^{2+}$-binding protein is described in volume 1, Chapter 10.
7. Ideally, all additions are equally large to get evenly spaced points. The aliquot has to be sufficiently small to get enough points in the titration for obtaining good precision in the binding constants. It is good to have the chelator covered by at least five points and each site in the protein by an additional five points, plus approx five points for the baseline after the binding is saturated. Aliquots of 3, 4, or 5 μL of 3 m*M* $CaCl_2$ are often ideal. If the measurements are performed at high salt and/or the protein or chelator binds calcium with lower affinity, one may need to add calcium from a stronger stock solution (e.g., 10 m*M*) at the end of the titration to get closer to saturation. Examples of curves for a lower affinity chelator (5N-BAPTA) and protein can be seen in **ref. *4***.
8. It may take time to reach equilibrium. One often has to make a compromise between the time it takes to obtain a stable recording and minimized photobleaching

of the chelator. Another problem with waiting too long is that the protein and/or chelator may start to absorb calcium form the cuvet. It is often best to wait 20–60 s until the initial quick changes in absorbance have settled and then note the recording. For a slowly equilibrating system, the cuvet may be put in darkness for equilibration for a few minutes at each titration point.

9. For example, if your absorbance is around 0.2 and you add 5-μL calcium aliquots to 2.5 mL titrand, the dilution will cause the absorbance to drop by 0.0004 at each addition.
10. An in-house computer program that performs the described analysis can be obtained from the author at Sara.Linse@Fkem2.LTH.SE.
11. The initial total calcium concentration in the titrand before any calcium additions are made can be determined by atomic absorption spectroscopy. Another way is to let the titration data specify the initial calcium concentration $Ca_0$ by fitting the data using several different values of $Ca_0$, and choosing the one that gives a value of (AMAX-AMIN) that agrees with $A_2$–$A_3$, (*see* **Subheading 3.1., step 3**).
12. It may be dangerous to use an adjustable factor *F* if the stoichiometry is not known or if the chelator and/or protein binds calcium with a lower affinity so that the curve does not have a sharp corner at the point of saturation. Using *F* to correct for protein concentration errors, however, works fine with a set of proteins with high affinity and known stoichiometry, because *F* is often better determined by the data than by any other method.
13. Such error bars do not include systematic errors because of false values of the calcium affinity for the chelator. Hence, they are a measure of precision rather than accuracy. Because the method is based on competition between the protein and a chelator, the precision is often high, and when the aim is to study the effects of mutations or other modifications, the method can be very useful and reliable.
14. The free energy of interaction between binding events at separate sites $\Delta\Delta G$ cannot be measured by the chelator method because this is a microscopic property. However, the macroscopic binding constants can be used to calculate a lower limit to $-\Delta\Delta G$. For a protein with two sites, this limiting value is $RT \ln (4K_2/K_1)$ and equal to the true cooperativity if the two sites have equal affinities. For a more thorough discussion of cooperativity and how it can be measured, *see* **refs. *5***, ***7***, and ***8***.
15. If the initial calcium concentration is not precisely known, or if it is too high (several μ*M*) the precision in the determined macroscopic-binding constants will be reduced. In such cases, it is especially difficult to quantitate the cooperativity, as points are missing in the beginning of the curve (*see* **Fig. 2C**). The initial curvature is not defined and the separation of the total affinity into individual macroscopic binding constants becomes uncertain.

## Acknowledgments

Helpful comments on the manuscript by Eva Thulin, Tord Berggård, and Jonas Fast are gratefully acknowledged.

## References


1. Tsien, R. Y. (1980) New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis and properties of prototype structures. *Biochemistry* **19,** 2396–2404.
2. Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmdén, K., et al. (1987) Structure-function relationships in EF-hand $Ca^{2+}$-binding proteins. Protein engineering and biophysical studies of calbindin $D_{9k}$. *Biochemistry* **26,** 6723–6735.
3. Haugland, R. (1996) Handbook of fluorescent probes and research chemicals. Molecular Probes, Inc., Eugene, Oregon.
4. Rand, M. D., Lindblom, A., Carlson, J, Villoutreix, B. O., and Stenflo, J. (1997) Calcium binding to tandem repeats of EGF-like modules. Expression and characterization of the EGF-like modules of human Notch-1 implicated in receptor-ligand interactions. *Protein Sci.* **6,** 2059–2071.
5. Linse, S., Sellers, P., and Thulin, E. (1993) Disulfide bonds in homo and heterodimers of EF-hand subdomains of calbindin $D_{9k}$: stability, calcium binding and NMR studies. *Protein Sci.* **2,** 985–1000.
6. Bylsma, N., Drakenberg, T., Andersson, I., Leadley, P. F., and Forsén, S. (1992) Prokaryotic calcium-binding protein of the calmodulin superfamily. Calcium binding to *Saccharopolyspora erythraea* 20 kDa protein. *FEBS Lett.* **299,** 44–47.
7. Linse, S., Johansson, C., Brodin, P., Grundström, T., Drakenberg, T., and Forsén, S. (1991). Electrostatic contribution to the binding of $Ca^{2+}$ in calbindin $D_{9k}$. *Biochemistry* **30,** 154–162.
8. Forsén, S. and Linse, S. (1995) Cooperativity: over the hill. *Tr. Biochem. Sci.* **20,** 495–497.

# 3

# Deconvolution of Calcium-Binding Curves

## *Facts and Fantasies*

**Jacques Haiech and Marie-Claude Kilhoffer**

## 1. Introduction

$Ca^{2+}$ signaling is of paramount importance in intracellular communication of eukaryotic cells. Many external stimuli trigger a transient change in the cytosolic-free $Ca^{2+}$ concentration (in the form of a $Ca^{2+}$ wave or $Ca^{2+}$ oscillations). The internal $Ca^{2+}$ modulation is deciphered by $Ca^{2+}$-binding proteins, which undergo conformational changes upon $Ca^{2+}$-binding allowing them to act as enzymatic or protein modulators. These $Ca^{2+}$-binding proteins have been well described in the past three decades ***(1–6)***. Calmodulin, an ubiquitous and multifunctional protein, is considered as the prototype of the $Ca^{2+}$-binding protein family containing EF-hand domains ***(7–16)***. Because of its pivotal role in many $Ca^{2+}$-dependent cellular events, the understanding of the mechanism of action of this protein at the molecular level has been the aim of several research groups. For such a study, four main points have to be tackled:

- Description of the mechanism of $Ca^{2+}$-binding to calmodulin;
- Understanding of the conformational changes induced by $Ca^{2+}$-binding;
- Analysis of the interaction of calmodulin with the different targets; and
- Deciphering the activation or the modulation of the calmodulin/target protein complexes.

Calmodulin is a protein with four $Ca^{2+}$-binding sites. In this chapter, we will describe how to deconvolute $Ca^{2+}$-binding curves, with special emphasis on calmodulin $Ca^{2+}$-binding curves.

The mechanism of $Ca^{2+}$-binding to calmodulin has been described since 1973 and has been subject to many controversies. Therefore, we are going

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel © Humana Press Inc., Totowa, NJ

to use the scientifical history of calmodulin to present different drawbacks (fantasies) that can happen when interpreting $Ca^{2+}$-binding data (facts).

When analyzing $Ca^{2+}$-binding or titration data, different authors used different definitions. Therefore, our first consideration will be to define the terms that we will use, even if, for sake of clarity, we do not always follow classical definitions. In general, we will stick to the logic of the book of I. M. Klotz ***(17)***

## 2. Macroscopic States vs Microscopic States

Let us first consider a protein with two sites for a given ligand ($Ca^{2+}$, for instance). The system may be represented in two ways (*see* **Fig. 1A**,**B**). The scheme on **Fig. 1A** will be termed macroscopic scheme and the scheme in **Fig. 1B**, the microscopic scheme.

From an intuitive point of view, the first scheme implies that with the experimental setup used, one cannot make a distinction between the two different sites of the proteins. Under these conditions, the system is described by three states (with the meaning used in thermodynamics, each state is characterized by state functions that describe the average properties of billions of particles):

- One without ligand on the protein (the apoform);
- One with one ligand on the protein (regardless of the site occupied); and
- One with two ligands (all the sites are occupied).

Consider that we measure the number of ligands bound per protein using equilibrium dialysis or flow dialysis. This experimental setup leads to binding data that can only be analyzed using the macroscopic scheme. Our aim is to be able to describe the ligand-binding mechanism at the molecular level. In other terms, we would like to use the second scheme and to describe the binding mechanism of each individual site. Although the average behavior of several millions or billions of molecules is being measured, our ultimate goal is not to describe the average state of the protein, but to visualize the behavior of each molecule upon $Ca^{2+}$-binding to each individual site.

When describing the binding mechanism in its macroscopic form (*see* **Fig. 1A**), two parameters $K_1$ and $K_2$, termed macroscopic-association constants, are used. To describe the same binding at the microscopic level, three parameters are used: $k_1$ and $k_2$, called microscopic- or individual-association constants, and $c$, a coupling factor (**Fig. 1B**). The coupling factor $c$ is always positive. If $c$ is equal to 1, the two sites are independent. Otherwise there is positive ($c > 1$) or negative ($c < 1$) coupling between the two sites. As the same molecular mechanism is described by two different schemes, there is a relationship between the macroscopic constants on the one hand, and the micro-

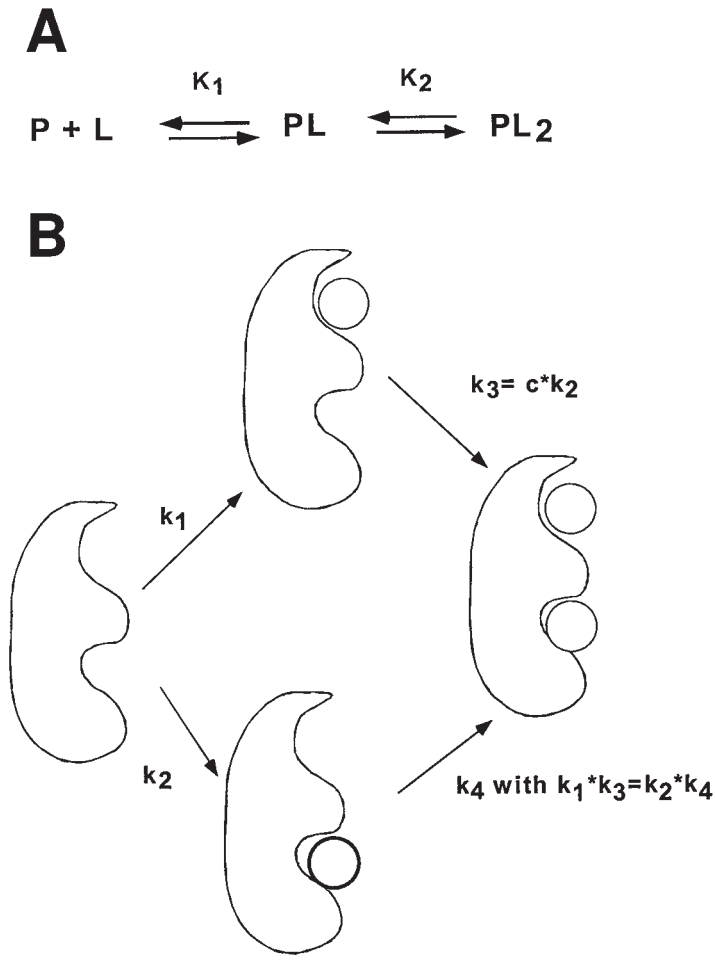


Fig. 1. Schematic representation of a protein with two binding sites for a specific ligand. **(A)** Macroscopic scheme; **(B)** Microscopic scheme.

scopic constants and the coupling factors on the other hand. From **Fig. 1**, it is straightforward to deduce the following:

$$K_1 = k_1 + k_2$$
$$K_1 * K_2 = c * k_1 * k_2 \quad (1)$$

Ligand-binding data obtained from equilibrium or flow-dialysis experiments are described by the so-called Adair-Klotz equation:

$$\gamma = K_1 * (L) + 2 * K_1 * K_2 (L)^2 / 1 + K_1 * (L) + K_1 * K_2 (L)^2 \quad (2)$$

where γ is the number of moles of ligand bound per mole of protein and (*L*) is the free-ligand concentration. The denominator of **Eq. 2** is the binding polynomial. The degree of this polynomial corresponds to the number of sites.

Nonlinear regression on the experimental data using **Eq. 2** allows to determine the two macroscopic constants. Determination of the degree of the binding polynomial is also possible. However, for obvious mathematical reasons, it would be much better to determine the number of sites for a given ligand using an independent technique (for instance, mass spectrometry). Determination of the macroscopic constants would then be much more precise.

From the macroscopic constants and using **Eq. 1**, we would like to determine the individual or microscopic constants and the coupling factor. Unfortunately, we have two equations with three unknowns. To interpret our "macroscopic data" in a "microscopic or molecular" scheme, we have to make some simplifying hypothesis. In others words, we have either to fix the value of at least one of the unknown parameters or to add a third equation.

Two hypotheses are classically found in the literature:

- One considers the sites to be independent; *c* is then equal to 1;
- The second uses a principle of symmetry and considers that the sites are identical; their individual association constants are then equal ($k_1 = k_2 = k$).

Using the first hypothesis, **Eq. 1** may be solved if

$$K_1 \geq 4 * K_2 \tag{3}$$

If this inequality does not hold, the system cannot be interpreted with *c* equal to 1. That implies that the two sites are not independent.

Using the second hypothesis, three equations with three unknowns are obtained. The solution of this system is

$$k_1 = k_2 = k = K_1/2 \tag{4}$$

Combining the two hypotheses, we assume that the two sites are independent and equivalent. This strong assumption is called the Scatchard hypothesis. With this assumption, **Eq. 1** becomes

$$\begin{aligned} K_1 &= 2 * k \\ K_1 * K_2 &= k^2 \end{aligned} \tag{5}$$

**Equation 5** can be solved if and only if

$$K_1 = 4 * K_2 = 2 * k \tag{6}$$

With this hypothesis, **Eq. 2** becomes

$$v = [2 * k * (L)] / [1 + k * (L)] \tag{7}$$

**Equation 7** may be rewritten

$$v/(L) = k * (2 - v) \tag{8}$$

This equation is called the equation of Scatchard and its graph is a straight line.

We have here an important result, which can be phrased as follows (the Scatchard theorem):

- A protein with two equivalent and independent site for a given ligand presents a ligand-binding curve, which is described by the Scatchard equation and its the graphical representation is a straight line.
- On the other hand, a protein-ligand binding curve, which is a straight line in the Scatchard representation does not imply that the protein has two independent and equivalent sites.

Indeed, when

$$K_1 = 4 * K_2 \tag{9}$$

the ligand-binding curve is a straight line in the Scatchard representation. We have an infinite number of possibilities to choose $k_1$, $k_2$, and $c$ to fulfil **Eq. 1** and **Eq. 9**. For example, for any value of $k_1$, the following triplet fulfills the previous requirements:

$$\begin{matrix} k_1 \\ k_2 = k_1 / 1000 \\ c = 250 \end{matrix} \tag{10}$$

In this example, although the Scatchard representation is a straight line, the sites are neither equivalent, nor independent. The protein has one site with high affinity for the ligand and one site with low affinity. Moreover, upon ligand binding to the high affinity site, there is a strong positive coupling with the second site.

In most textbooks, authors consider that when ligand-binding data of a protein lead to a straight line in the Scatchard representation, the protein has equivalent and independent sites for the ligand. This explains several controversies in the $Ca^{2+}$-binding protein field. This explains several controversies in the $Ca^{2+}$-binding protein field.

## 3. A General Model for Whatever the Signal Is

From the previous reasoning, we are faced with the following problem: the $Ca^{2+}$-binding curve alone does not allow to provide a single and genuine molecular-binding mechanism. In this context, experiments that allow us to track each individual complex of the scheme in **Fig. 1B** would be extremely useful. Several research groups aimed to develop and introduce in the protein

reporter groups with spectroscopic properties that are sensitive to the occupancy of one specific site by the ligand.

Assume that to each complex depicted in **Fig. 1B**, we associate a signal, namely $s_0$ for the protein without ligand, $s_1$ for the protein with ligand in site 1, $s_2$ for the protein with ligand in site 2, and $s_3$ for the protein with two ligands. It is straightforward to derive the equation describing the variation of the signal as a function of the free-ligand concentration ($L$).

$$S = \frac{s_0 + s_1 * k_1 * (L) + s_2 * k_2 * (L) + s_3 * c * k_1 * k_2 * (L)^2}{1 + k_1 * (L) + k_2 * (L) + c * k_1 * k_2 * (L)^2} \tag{11}$$

Notice that $s = s_0$ when $(L) = 0$ and $s = s_3$ for $(L)$ at saturating concentration. This last point means that from an experimental point of view, we may increase the concentration of $(L)$ in such a way that $c * k_1 * k_2 * (L)^2 >> 1 + (k_1 + k_2) * (L)$. As the concentration of $(L)$ may be limited to a given range, the unknown $s_3$ cannot be always determined independently.

If in a given experiment, the signal $S$ we measure corresponds to the number of ligand bound per protein, we have

$$s_0 = 0;\ s_1 = s_2 = 1$$

and $s_3 = 2$ (the signal corresponds to the number of bound ligands for each complex).

**Equation 11** can be rewritten:

$$v = \frac{k_1 * (L) + k_2 * (L) + 2 * c * k_1 * k_2 * (L)^2}{1 + k_1 * (L) + k_2 * (L) + c * k_1 * k_2 * (L)^2} \tag{12}$$

This equation is equivalent to **Eq. 2** combined with **Eq. 1**.

Assume now that we are able to introduce, at a specific location in the protein, a reporter group sensitive to the occupancy of the site 1 (respectively, a reporter group sensitive to the occupancy of site 2). Therefore, for the signal arising from the first reporter group, we have:

$$S^* = \frac{k_1 * (L) + c * k_1 * k_2 * (L)^2}{1 + k_1 * (L) + k_2 * (L) + c * k_1 * k_2 * (L)^2} \tag{13}$$

as $s_0 = 0$, $s_1 = 1$, $s_2 = 0$, and $s_3 = 1$ (in relative units).

For the second reporter group, sensitive to the occupancy of site 2, we have:

$$S^{**} = \frac{k_2 * (L) + c * k_1 * k_2 * (L)^2}{1 + k_1 * (L) + k_2 * (L) + c * k_1 * k_2 * (L)^2} \tag{14}$$

as $s_0 = 0$, $s_1 = 0$, $s_2 = 1$, and $s_3 = 1$ (in relative units).

**Figure 2** presents a graphical representation of **Eqs. 12–14** in the case of a protein with two independent and equivalent sites (*see* **Fig. 2A**) and in the case

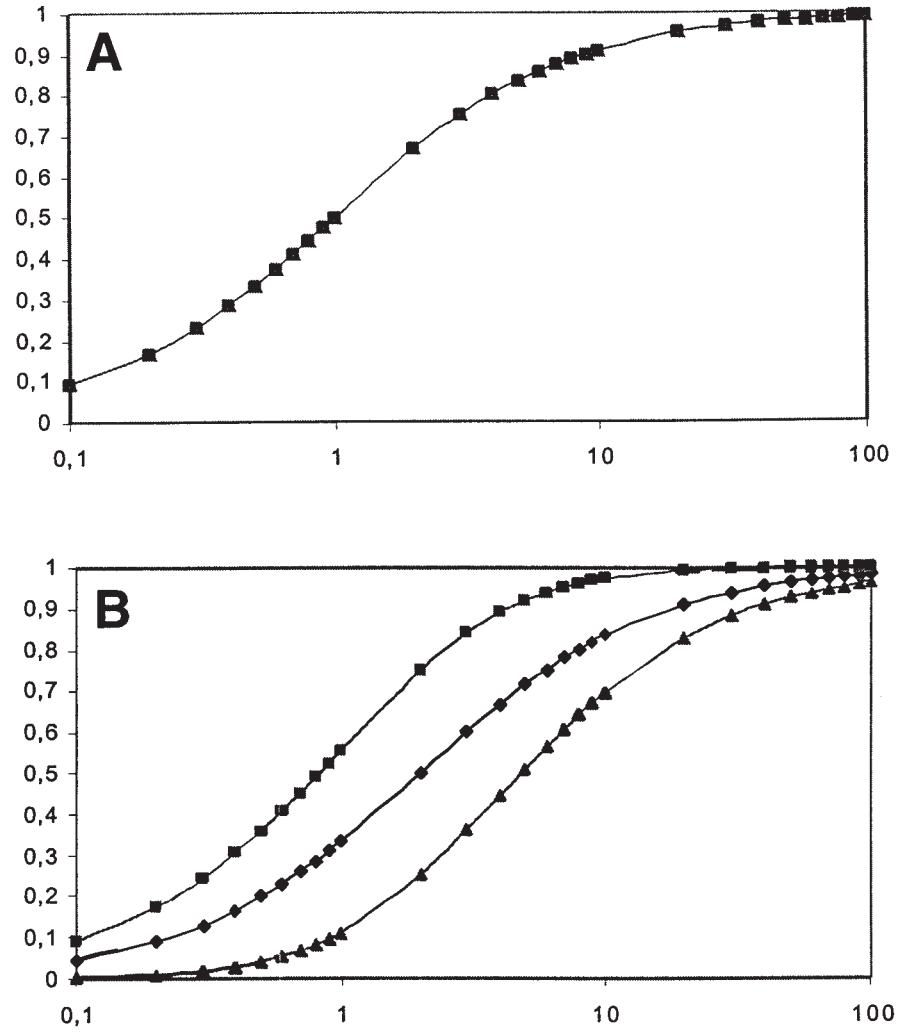


Fig. 2. Graphical representation of **Eqs. 12–14** with $k_1 = k_2 = c = 1$ **(A)** and $k_1 = 1$, $k_2 = 0{,}001$, and $c = 250$ **(B)**. *X*-axis represents the normalized concentration of the ligand and *y*-axis, the normalized amplitude of the signal ($s_1$ ■, $v$ ◆, and $s_2$ ▲).

of a protein with a high-affinity and a low-affinity site with strong positive coupling between the two (*see* **Fig. 2B**).

In the first case, all three curves are identical. In the second case, the $Ca^{2+}$-binding curve is akin to the previous one, but the curves associated with the occupancy of site 1 and site 2 are, respectively, left-shifted and right-shifted. In this latter model, upon ligand titration, the first site is occupied before the second site. However, the mean number of bound ligand is the same as for a pro-

tein with independent and equivalent sites. We will call this binding mechanism a sequential binding mechanism.

At this point, we want to underline that for a protein with more than one site for a ligand *L*, the ligand-binding curve obtained by flow or equilibrium dialysis and the spectroscopic ligand-binding curves have to be combined in order to refine the possible molecular interpretations and cannot be used independently from one another.

The previous equations can easily be generalized to proteins with *n* sites for a given ligand and even for several different ligands. For a protein with four sites (such as calmodulin which binds four $Ca^{2+}$), we have to deal with four macroscopic constants, four microscopic or individual constants, and 11 coupling factors.

**Equations 2**, **11**, and **12** take the following form:

$$v = \frac{K_1 * (L) + 2 * K_1 * K_2 * (L)^2 + 3 * K_1 * K_2 * K_3 * (L)^3 + 4 * K_1 * K_2 * K_3 * K_4 (L)^4}{1 + K_1 * (L) + K_1 * ZK_2 * (L)^2 + K_1 * K_2 * K_3 * (L)^3 + K_1 * K_2 * K_3 * K_4 (L)^4} \tag{15}$$

$$S = \frac{S_0 + (\sum_{1}^{4} s_i k_i) * (L) + (\sum_{i<j}^{1 \le (i,j) \le 4} s_{i,j} * c_{i,j} * k_i * k_j) * (L)^2 + (\sum_{i<j<k}^{1 \le (i,j,k) \le 4} s_{i,j,k} * c_{i,j,k} * k_i * k_j * k_k) * (L)^3 + s_{1,2,3,4} * c_{1,2,3,4} * k_1 * k_2 * k_3 * k_4 * (L)^4}{1 + (\sum_{1}^{4} k_i) * (L) + (\sum_{i<j}^{1 \le (i,j) \le 4} c_{i,j} * k_i * k_j) * (L)^2 + (\sum_{i<j<k}^{1 \le (i,j,k) \le 4} c_{i,j,k} * k_i * k_j * k_k) * (L)^3 + c_{1,2,3,4} * k_1 * k_2 * k_3 * k_4 * (L)^4} \tag{16}$$

$$v = \frac{(\sum_{1}^{4} k_i) * (L) + 2 * (\sum_{i<j}^{1 \le (i,j) \le 4} c_{i,j} * k_i * k_j) * (L)^2 + 3 * (\sum_{i<j<k}^{1 \le (i,j,k) \le 4} c_{i,j,k} * k_i * k_j * k_k) * (L)^3 + 4 * c_{1,2,3,4} * k_1 * k_2 * k_3 * k_4 * (L)^4}{1 + (\sum_{1}^{4} k_i) * (L) + (\sum_{i<j}^{1 \le (i,j) \le 4} c_{i,j} * k_i * k_j) * (L)^2 + (\sum_{i<j<k}^{1 \le (i,j,k) \le 4} c_{i,j,k} * k_i * k_j * k_k) * (L)^3 + c_{1,2,3,4} * k_1 * k_2 * k_3 * k_4 * (L)^4} \tag{17}$$

The remarks stressed out for a protein with two sites, also apply to a protein with *n* sites ($n > 1$).

## 4. Analysis of Ligand-Binding Curves

To analyze the ligand binding to a protein, we follow a specific signal as a function of the free-ligand concentration (titration) or total-ligand concentration (addition). The free-ligand concentration may be directly measured as in equilibrium and flow-dialysis experiments, or using a ligand-sensitive electrode. It also may be clamped by a ligand-buffering system. In most spectrometric experiments, we have access to the total added ligand concentration and we know the protein concentration in our cuvet.

In titration experiments, we determine a set of couples $(S_i,(L)_i)$ where *S* is the signal as defined above, $(L)$ the free-ligand concentration, and *i* varies between 1 and *n* (*n* being the total number of experimental points). For a pro-

tein with four binding sites for ligand $L$, we have to find by nonlinear regression the best parameters in order to fit the experimental points to **Eq. 17**. Using dialysis (equilibrium or flow) and spectroscopic experiments, we may combine **Eqs. 16** and **17**.

In the experiments were addition is performed, we have to solve a nonlinear system of three equations:

$$v = \frac{(\sum_{1}^{4} k_i)*(L)+2*(\sum_{i<j}^{1\le(i,j)\le 4} c_{i,j}*k_i*k_j)*(L)^2+3*(\sum_{i<j<k}^{1\le(i,j,k)\le 4} c_{i,j,k}*k_i*k_j*k_k)*(L)^3+4*c_{1,2,3,4}*k_1*k_2*k_3*k_4*(L)^4}{1+(\sum_{1}^{4} k_i)*(L)+(\sum_{i<j}^{1\le(i,j)\le 4} c_{i,j}*k_i*k_j)*(L)^2+(\sum_{i<j<k}^{1\le(i,j,k)\le 4} c_{i,j,k}*k_i*k_j*k_k)*(L)^3+c_{1,2,3,4}*k_1*k_2*k_3*k_4*(L)^4} \quad (17)$$

$$S = \frac{S_0+(\sum_{1}^{4} s_i k_i)*(L)+(\sum_{i<j}^{1\le(i,j)\le 4} s_{i,j}*c_{i,j}*k_i*k_j)*(L)^2+(\sum_{i<j<k}^{1\le(i,j,k)\le 4} s_{i,j,k}*c_{i,j,k}*k_i*k_j*k_k)*(L)^3+s_{1,2,3,4}*c_{1,2,3,4}*k_1*k_2*k_3*k_4*(L)^4}{1+(\sum_{1}^{4} k_i)*(L)+(\sum_{i<j}^{1\le(i,j)\le 4} c_{i,j}*k_i*k_j)*(L)^2+(\sum_{i<j<k}^{1\le(i,j,k)\le 4} c_{i,j,k}*k_i*k_j*k_k)*(L)^3+c_{1,2,3,4}*k_1*k_2*k_3*k_4*(L)^4} \quad (16)$$

$$L_t = v * P_t + (L)$$

where $P_t$ is the total concentration of protein, $L_t$ is the total concentration of ligand.

Such a system of equations can be solved by nonlinear regression. The determination of $v$, the number of sites, by an independent experiment is greatly recommended as it facilitates the resolution of the equations.

Software such as Excel, Origin, or SAS may handle easily such systems of equations.

Although few work has been done in this field, it would be interesting to develop statistical tests that are pertaining to this kind of mathematical equations or systems of equations.

## 5. Deciphering the Mechanism of $Ca^{2+}$ Binding to Calmodulin

Since 1970, the intracellular eukaryotic $Ca^{2+}$-binding proteins appear as the main $Ca^{2+}$ detectors ***(1–3,5,6,18)***. Most of these proteins belong to the EF-hand domain protein family and present in their structure the canonical EF-hand domain, constituted by two 12 residue alpha helices surrounding a 12 residue $Ca^{2+}$-binding loop ***(4,19–21)***, suggesting their probable evolution from a single EF-hand domain by duplication. The prototype of this family is calmodulin, a four EF-hand domain protein ***(8,10,12–16)***. Whereas most of the $Ca^{2+}$-binding proteins are specifically localized and are representative of a given cellular state, calmodulin appears to be ubiquitous, present in all eukaryotic species, and involved in a multitude of $Ca^{2+}$-dependent cellular events, through its interaction with various target enzymes. Therefore, numerous studies were undertaken in order to get detailed mechanistic insight into $Ca^{2+}$-binding to this fascinating protein.

Calmodulin was identified as an activator of cyclic nucleotide phosphodiesterase in 1970 by Cheung and Kakiuchi ***(7,9)***. The biological activity of the protein was investigated during the 1970s ***(22–30)***, but the complete amino acid sequence appeared only in 1980 ***(31,32)***. Calmodulin crystallization was difficult and the first 3D-structure was released in 1985 and refined in 1988 ***(33,34)***. In its crystal form, the protein appeared as a dumbbell, composed of two lobes linked by a long alpha helix. Each lobe is composed of two EF-hand $Ca^{2+}$-binding domains.

Deciphering the $Ca^{2+}$-binding mechanism of calmodulin was some kind of challenge for several teams since 1973.

### 5.1. Evolution of the Calmodulin models

In the 1970s, starting with Teo and Wang ***(35)***, several groups ***(36–44)*** performed $Ca^{2+}$ titrations of calmodulin using equilibrium or flow dialysis. Most of the experimental data were represented using the Scatchard representation. This representation of the data led to a straight line and pointed to the presence of three to four $Ca^{2+}$-binding sites, depending upon the experimental conditions. A consensus rapidly appeared in the calmodulin field, which considered calmodulin as a protein with four independent and equivalent sites. At high-ionic strength, one of the sites was weakened and only three sites were titratable.

Then at the beginning of the 1980s, NMR was used to follow the conformational changes of calmodulin upon $Ca^{2+}$ addition and the $Ca^{2+}$-binding sites were characterized using $^{43}Ca$ and $^{113}Cd$ ***(45–50)***. A new consensus emerged in which calmodulin was considered as a protein with four independent sites divided into two classes: two high-affinity sites in the C-terminal lobe and two low-affinity sites in the N-terminal lobe. Results from several other spectroscopic techniques (UV and fluorescence spectroscopy, circular dichroism ***[51–55]***) fitted well to this last model. This was the start of a controversy as the two models are exclusive one from the other. Although we published in 1981 ***(38)*** an interpretation that allowed to reconcile the different experimental data (based on the model in **Fig. 2A** but applied to a protein with 4 sites), such interpretation was ignored. In 1985, Wang ***(56)*** proposed a model to reconcile linear Scatchard plots and spectroscopic data. The properties of calmodulin in this model were the following:

- Calmodulin has two independent pairs of sites;
- Each pair of sites exhibits positive cooperativity on $Ca^{2+}$-binding; and
- The pair of sites in the COOH terminal half binds $Ca^{2+}$ with higher affinity than the pair of sites in the $NH_2$ terminal half.

This model uses three parameters: two individual constants (one for each pair of sites) and one coupling factor (the same for each pair). With a good choice

of the parameters, experimental data obtained, either by measuring the number of ligands bound to the protein or by following a spectroscopic signal as a function of the ligand concentration, can be fitted to the model. One of the main assumption of the model of Wang is the independence of the two lobes. If this assumption is true, the behavior of whole calmodulin must be the sum of the behaviors of its individual lobes (which can be obtained by proteolytic digestion ***[51,57]***). Some studies concluded that there was little or no interaction between the *N*-terminal and *C*-terminal halves of calmodulin. This was based on comparisons of titration curves of whole calmodulin *vs* those of isolated domains ***(58–60)***. Other studies, using mutated calmodulins, were in favor of an interaction between the two domains ***(61–64)***. A controversy arose on this issue, as one cannot be sure that the tryptic fragments generated retain their original structure, nor that mutagenesis does not induce minor changes in the protein structure. It is only recently that work performed on whole calmodulin set up his controversy and clearly established that the two lobes of calmodulin were not independent ***(65–68)***. Also, in the presence of target proteins or peptides, the cooperativity between the two lobes of CaM markedly increases.

To get more insight into the mechanism of $Ca^{2+}$-binding to calmodulin, we decided in 1986 to build isofunctional mutants with an internal reporter group ***(64,69,70)*** that will give us access to the occupancy of one specific site of the protein. Our strategy was based on the reasoning developed in **Subheading 3.** (*see* **Fig. 2**). As standard calmodulin does not harbor tryptophan residues in its structure, we introduced one tryptophanyl residue at specific positions in calmodulin ***(64)***. Results allowed us to confirm the model we proposed in 1981. This model is presented in **Fig. 3**.

Each site exists in two different conformations: a conformation of high affinity for $Ca^{2+}$ (HC conformation) and a low-affinity $Ca^{2+}$-binding conformation (LC conformation). In the apoform, only site III is in the HC conformation. Upon $Ca^{2+}$ occupancy of this site, site IV undergoes a conformational change toward a HC conformation, and so on, step by step, for the remaining sites. This model is basically described by four constants, the individual binding constants of each individual site in the HC conformation. If $g1$, $g2$, $g3$, and $g4$ are these constants, the equation describing $Ca^{2+}$-binding to calmodulin can be written

$$v = \frac{g_1 * (L) + 2 * g_1 * g_2 * (L)^2 + 3 * g_1 * g_2 * g_3 * (L)^3 + 4 * g_1 * g_2 * g_3 * g_4 (L)^4}{1 + g_1 * (L) + g_1 * g_2 * (L)^2 + g_1 * g_2 * g_3 * (L)^3 + g_1 * g_2 * g_3 * g_4 (L)^4} \quad (18)$$

These constants ($g_i$) correspond to the macroscopic constants of the Adair-Klotz **Eq. 2**. In this model, we have a direct interpretation of the macroscopic constants in terms of individual constants and coupling factors.

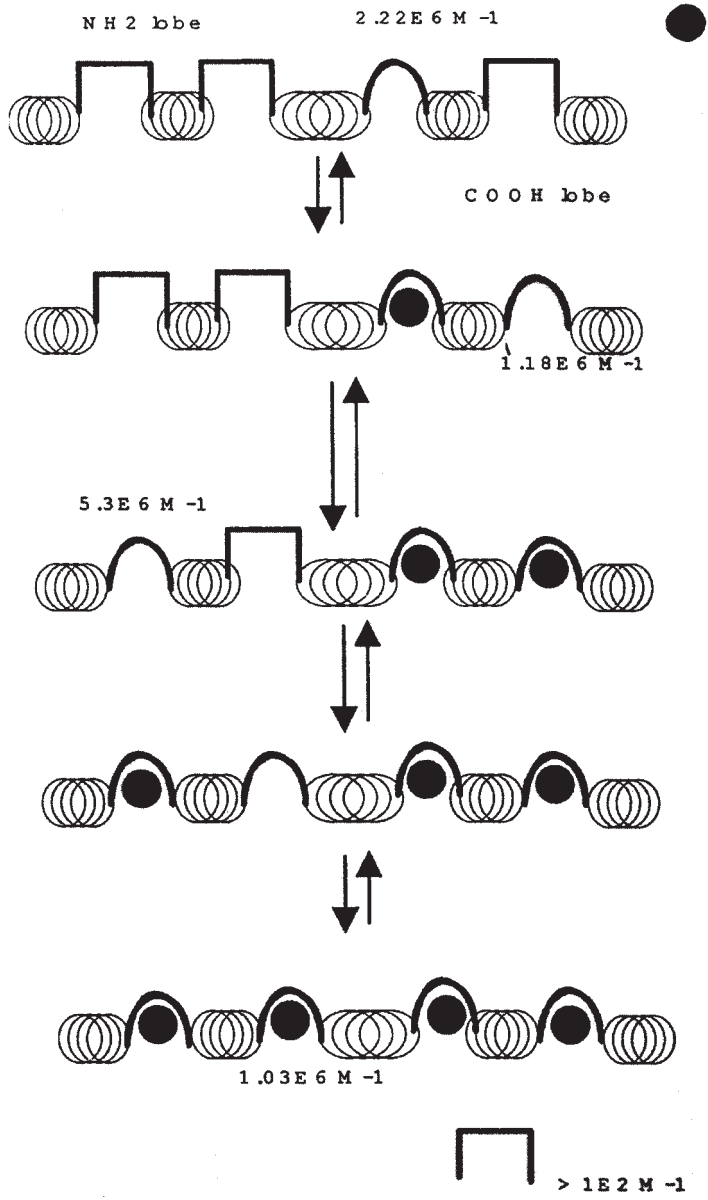


Fig. 3. Mechanism of $Ca^{2+}$-binding to calmodulin. The big circles correspond to $Ca^{2+}$, the thether to the helices and connecting peptides. At the start, all sites but site III, are in a conformation of low affinity for $Ca^{2+}$ (LC conformation). After each $Ca^{2+}$-binding step, one specific site acquires the high-$Ca^{2+}$-affinity conformation (HC conformation). Numbers correspond to the $Ca^{2+}$-affinity constants.

As calmodulin belongs to the EF-hand family of intracellular $Ca^{2+}$-binding proteins, we may wonder if this binding mechanism is pertinent to most proteins of the family. Moreover, the number of $Ca^{2+}$-binding sites varies from 2 (parvalbumin) to 6 (for review, *see* **ref. *18***). In 1979, we already showed that the same general model applies to parvalbumin ***(71)***. As parvalbumin has two EF-hand $Ca^{2+}$-binding sites, the parameters of such a model were easier to describe. Parvalbumin is not very close to calmodulin from an evolutionary point of view. Therefore, a principle of simplicity leads us to suggest that the sequential binding model is probably pertinent to most EF-hand $Ca^{2+}$-binding proteins.

## 6. Open questions

Recent studies on calmodulin using mass spectrometry have shown that calmodulin exhibits more than four binding sites for $Ca^{2+}$ and for others cations ***(72)***. Other $Ca^{2+}$-binding proteins exhibit cationic binding sites ($Zn^{2+}$ or $Cu^{2+}$ for S100 ***[73]***, a third site for Parvalbumin ***[74]***). This suggests that, in addition to the so called EF $Ca^{2+}$-binding sites, others cationic sites exist on these proteins. These sites may bind $Ca^{2+}$ and/or $Mg^{2+}$ among others cations with low specificity. However, up to now, there are only few studies on the physiological role of these sites. Recently, using microcalorimetry, we have shown that $Mg^{2+}$ acts through these cationic sites to modulate the $Ca^{2+}$-binding affinities of the individual EF hand sites of calmodulin. We also brought evidence that $Mg^{2+}$ modulates the interaction between the two lobes of calmodulin ***(75)***.

Other experimental conditions (ionic strength, presence of target structures or peptides such as the RS20 or M13 [the binding domain of myosin light chain kinase of smooth muscle or skeletal muscle, respectively], others cations, ethylene glycol) modify the $Ca^{2+}$-binding parameters of the system, but qualitatively the mechanism of $Ca^{2+}$-binding to calmodulin seems to remain unchanged. In the cell, the water concentration is likely to be different from the one prevailing in the test tube. The observation that ethylene glycol (which decreases the water concentration) has a strong effect on calmodulin ***(64)***, suggests that depending on the localization of calmodulin within the cell, its $Ca^{2+}$-binding parameters and, therefore, its $Ca^{2+}$-binding kinetics may be different. Monitoring the different calmodulin/$Ca^{2+}$ complexes directly in the various cellular compartment seems now to be of paramount importance if one has to get deeper insight into how calmodulin plays its multifunctional role in the cell.

## 7. Conclusions

Most of the controversies in the $Ca^{2+}$-binding protein field have arisen from a misuse or misunderstanding of the equations governing the interaction between a protein and a ligand, when the protein is a multisite protein.

We may suggest the following conjectures:

- The set of binding equations probably has mathematical properties that it would be interesting to decipher in order to improve the statistical tests associated with the deconvolution of calcium binding experimental curves;
- The use of several experimental techniques (namely, mass spectrometry, equilibrium or flow dialysis, microcalorimetry, spectroscopic techniques, and so on) is needed to cross the possible molecular interpretations; and
- With the progress in molecular dynamics, the molecular interpretation of a protein system has to be simulated.

New emerging techniques let envisage that in the coming years, the interaction between a protein and its ligands will be performed directly inside the cell and at the level of the single molecule. Such a fantastic development will shed new light on how cells manage $Ca^{2+}$ signaling.

## References

1. Gerke, V., Heizmann, C. W., and Krebs J. (1998) Special Issue. 5th European symposium on calcium binding proteins in normal and transformed cells, in *Biochem. Biophys. Acta. Molecular Cell Research*, vol. 1148 (2) (Avruch, J., ed.), Elsevier, Amsterdam.
2. Wasserman, R. H., Corradino, R. A., Carafoli, E., Kretsinger, R. H., MacLennan, D. H., and Siegel, F. L. (1977) *Calcium — Binding Proteins and Calcium Function*, North Holland, New York.
3. Heizmann, C. W. (1991) Novel calcium binding proteins, in *Fundamentals and Clinical Implications*, Springer-Verlag, Berlin, Heidelberg.
4. Kretsinger, R. H. (1975) Hypothesis: calcium modulated proteins contain EF hands, in *Calcium Transport in Contraction and Secretion* (Carafoli, E., et al., eds.), Amsterdam, North-Holland, pp. 469–78.
5. Pochet, R., Lawson, D. E. M., and Heizmann, C. W. (1989) Calcium binding proteins in normal and transformed cells, in *Advances in Experimental Medicine and Biology*, vol. 269, Plenum, New York.
6. Goodman, M., Pechere, J. F., Haiech, J., and Demaille, J. G. (1979) Evolutionary diversification of structure and function in the family of intracellular calcium-binding proteins. *J. Mol. Evol.* **13,** 331–352.
7. Cheung, W. Y. (1970) Cyclic 3',5'-nucleotide phosphodiesterase. Demonstration of an activator. *Biochem. Biophys. Res. Commun.* **38,** 533–538.
8. Cheung, W. Y. (1980) Calmodulin plays a pivotal role in cellular regulation. *Science* **207,** 19–27.
9. Kakiuchi, S. and Yamazaki, R. (1970) Calcium dependent phosphodiesterase activity and its activating factor (PAF) from brain studies on cyclic 3',5'-nucleotide phosphodiesterase (3). *Biochem. Biophys. Res. Commun.* **41,** 1104–1110.
10. Klee, C. B., Crouch, T. H., and Richman, P. G. (1980) Calmodulin. *Annu. Rev. Biochem.* **49,** 489–515.
11. Klee, C. B. and Vanaman, T. C. (1982) Calmodulin. *Adv. Prot. Chem*. **35,** 213–321.

12. Watterson, D. M. (1980) *Calmodulin and Cell Functions*, vol. 356, The New York Academy of Sciences, New York.
13. Marx, J. L. (1980) Calmodulin: a protein for all seasons. *Science* **208,** 274–276.
14. Cox, J. A., Comte, M., Malnoe, A., and Stein, E. A. (1984) Mode of action of the regulatory protein calmodulin, in *Metal Ions in Biological Systems*, vol. 17, (Sigel, H., ed.), Marcel Dekker, New York, pp. 215–273.
15. Watterson, D. M., Burgess, W. H., Lukas, T. J., Iverson, D., Marshak, D. R., Schleicher, M., et al. (1984) Towards a molecular and atomic anatomy of calmodulin and calmodulin-binding proteins, in *Advances in Cyclic Nucleotide and Protein Phosphorylation Research* (Strada, S. J., ed.), Raven, New York, pp. 205–226.
16. Cohen, P. (1988) Calmodulin, in *Molecular Aspects of Cellular Regulation*, vol. 5 (Cohen, P., ed.), Elsevier, Amsterdam.
17. Klotz, I. M. (1997) *Ligand-Receptor Energetics: A Guide for the Perplexed*, Wiley, New York.
18. Moncrief, N. D., Kretsinger, R. H., and Goodman, M. (1990) Evolution of EF-hand calcium-modulated proteins. I. Relationships based on amino acid sequences. *J. Mol. Evol.* **30,** 522–562.
19. Kretsinger, R. H. and Nockolds, C. E. (1973) Carp muscle calcium-binding protein. II. Structure determination and general description. *J. Biol. Chem.* **248,** 3313–3326.
20. Kretsinger, R. H. (1979) The informational role of calcium in the cytosol. *Adv. Cyclic Nucleotide Res*. **11,** 1–26.
21. Kawasaki, H. and Kretsinger, R. H. (1995) Calcium-binding proteins 1: EF-hands. *Protein Profile* **2,** 297–490.
22. Anderson, J. M. and Cormier, M. J. (1978) Calcium-dependent regulation of NAD kinase. *Biochem. Biophys. Res. Commun.* **84,** 595–602.
23. Brostrom, C. O., Huang, Y. C., Breckenridge, B. M., and Wolff, D. J. (1975) Identification of a calcium-binding protein as a calcium-dependent regulator of brain adenylate cyclase. *Proc. Natl. Acad. Sci. USA* **72,** 64–68.
24. Cohen, P., Burchell, A., Foulkes, J. G., and Cohen, P. T. (1978) Identification of the $Ca^{2+}$-dependent modulator protein as the fourth subunit of rabbit skeletal muscle phosphorylase kinase. *FEBS Lett.* **92,** 287–293.
25. Dabrowska, R., Aromatorio, D., Sherry, J. M., and Hartshorne, D. J. (1977) Composition of the myosin light chain kinase from chicken gizzard. *Biochem. Biophys. Res. Commun.* **78,** 1263–1272.
26. Gopinath, R. M. and Vincenzi, F. F. (1977) Phosphodiesterase protein activator mimics red blood cell cytoplasmic activator of ($Ca^{2+}$-$Mg^{2+}$)ATPase. *Biochem. Biophys. Res. Commun.* **77,** 1203–1209.
27. Jarrett, H. W. and Penniston, J. T. (1977) Partial purification of the $Ca^{2+}$-$Mg^{2+}$ ATPase activator from human erythrocytes: its similarity to the activator of 3':5'- cyclic nucleotide phosphodiesterase. *Biochem. Biophys. Res. Commun.* **77,** 1210–1216.
28. Schulman, H. and Greengard, P. (1978) Stimulation of brain membrane protein phosphorylation by calcium and an endogenous heat-stable protein. *Nature* **271,** 478–479.

29. Stewart, A. A., Ingebritsen, T. S., Manalan, A., Klee, C. B., and Cohen, P. (1982) Discovery of a $Ca^{2+}$- and calmodulin-dependent protein phosphatase: probable identity with calcineurin (CaM-BP80). *FEBS Lett.* **137,** 80–84.
30. Yazawa, M. and Yagi, K. (1977) Calcium-binding subunit of myosin light chain kinase. *J. Biochem. (Tokyo)* **82,** 287–289.
31. Watterson, D. M., Sharief, F., and Vanaman, T. C. (1980) The complete amino acid sequence of the $Ca^{2+}$-dependent modulator protein (calmodulin) of bovine brain. *J. Biol. Chem.* **255,** 962–975.
32. Watterson, D. M., Iverson, D. B., and Van Eldik, L. J. (1980) Spinach calmodulin: isolation, characterization, and comparison with vertebrate calmodulins. *Biochemistry* **19,** 5762–5768.
33. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R., and Cook, W. J. (1985) Three-dimensional structure of calmodulin. *Nature* **315,** 37–40.
34. Babu, Y. S., Bugg, C. E., and Cook, W. J. (1988) Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.* **204,** 191–204.
35. Teo, T. S. and Wang, J. H. (1973) Mechanism of activation of a cyclic adenosine 3':5'-monophosphate phosphodiesterase from bovine heart by calcium ions. Identification of the protein activator as a $Ca^{2+}$ binding protein. *J. Biol. Chem.* **248,** 5950–5955.
36. Crouch, T. H. and Klee, C. B. (1980) Positive cooperative binding of calcium to bovine brain calmodulin. *Biochemistry* **19,** 3692–3698.
37. Dedman, J. R., Potter, J. D., Jackson, R. L., Johnson, J. D., and Means, A. R. (1977) Physicochemical properties of rat testis $Ca^{2+}$-dependent regulator protein of cyclic nucleotide phosphodiesterase. Relationship of $Ca^{2+}$-binding, conformational changes, and phosphodiesterase activity. *J. Biol. Chem.* **252,** 8415–8422.
38. Haiech, J., Klee, C. B., and Demaille, J. G. (1981) Effects of cations on affinity of calmodulin for calcium: ordered binding of calcium ions allows the specific activation of calmodulin-stimulated enzymes. *Biochemistry* **20,** 3890–3897.
39. Jarrett, H. W. and Kyte, J. (1979) Human erythrocyte calmodulin. Further chemical characterization and the site of its interaction with the membrane. *J. Biol. Chem.* **254,** 8237–8244.
40. Keller, C. H., Olwin, B. B., LaPorte, D. C., and Storm, D. R. (1982) Determination of the free-energy coupling for binding of calcium ions and troponin I to calmodulin. *Biochemistry* **21,** 156–162.
41. Wolff, D. J., Poirier, P. G., Brostrom, C. O., and Brostrom, M. A. (1977) Divalent cation binding properties of bovine brain $Ca^{2+}$-dependent regulator protein. *J. Biol. Chem.* **252,** 4108–4117.
42. Burger, D., Cox, J. A., Comte, M., and Stein, E. A. (1984) Sequential conformational changes in calmodulin upon binding of calcium. *Biochemistry* **23,** 1966–1971.
43. Huang, C. Y., Chau, V., Chock, P. B., Wang, J. H., and Sharma, R. K. (1981) Mechanism of activation of cyclic nucleotide phosphodiesterase: requirement of the binding of four Ca2+ to calmodulin for activation. *Proc. Natl. Acad. Sci. USA* **78,** 871–874.
44. Ogawa, Y. and Tanokura, M. (1984) Calcium binding to calmodulin: effects of ionic strength, $Mg^{2+}$, pH and temperature. *J. Biochem. (Tokyo)* **95,** 19–28.

45. Forsen, S., Thulin, E., Drakenberg, T., Krebs, J., and Seamon, K. (1980) A 113Cd NMR study of calmodulin and its interaction with calcium, magnesium and trifluoperazine. *FEBS Lett.* **117,** 189–194.
46. Seamon, K. (1980) NMR studies on tyrosine-138 of calmodulin. *Ann. NY Acad. Sci.* **356,** 433–434.
47. Krebs, J., Carafoli, E., Thulin, E., and Forsen, S. (1980) 1H- and 113Cd-NMR studies of calmodulin. *Ann. NY Acad. Sci.* **356,** 397–398.
48. Andersson, T., Drakenberg, T., Forsen, S., and Thulin, E. (1982) Characterization of the $Ca^{2+}$ binding sites of calmodulin from bovine testis using 43Ca and 113Cd NMR. *Eur. J. Biochem.* **126,** 501–505.
49. Teleman, A., Drakenberg, T., and Forsen, S. (1986) Kinetics of $Ca^{2+}$ binding to calmodulin and its tryptic fragments studied by 43Ca-NMR. *Biochim. Biophys. Acta.* **873,** 204–213.
50. Yazawa, M., Kawamura, E., Minowa, O., Yagi, K., Ikura, M., and Hikichi, K. (1984) N-terminal region (domain 1) of calmodulin is the low affinity site for $Ca^{2+}$. A 13C NMR study of S-cyanocalmodulin. *J. Biochem. (Tokyo)* **95,** 443–446.
51. Drabikowski, W., Kuznicki, J., and Grabarek, Z. (1977) Similarity in Ca2+-induced changes between troponic-C and protein activator of 3':5'-cyclic nucleotide phosphodiesterase and their tryptic fragments. *Biochim. Biophys. Acta.* **485,** 124–133.
52. Klee, C. B. (1977) Conformational transition accompanying the binding of $Ca^{2+}$ to the protein activator of 3',5'-cyclic adenosine monophosphate phosphodiesterase. *Biochemistry* **16,** 1017–1024.
53. Richman, P. G. and Klee, C. B. (1979) Specific perturbation by $Ca^{2+}$ of tyrosyl residue 138 of calmodulin. *J. Biol. Chem.* **254,** 5372–5376.
54. Drabikowski, W., Brzeska, H., Kuznicki, J., and Grabarek, Z. (1980) Studies on structure and function of calmodulin. *Ann. NY Acad. Sci.* **356,** 374–375.
55. Kilhoffer, M. C., Demaille, J. G., and Gerard, D. (1981) Tyrosine fluorescence of ram testis and octopus calmodulins. Effects of calcium, magnesium, and ionic strength. *Biochemistry* **20,** 4407–4414.
56. Wang, C. L. (1985) A note on $Ca^{2+}$ binding to calmodulin. *Biochem. Biophys. Res. Commun.* **130,** 426–430.
57. Drabikowski, W., Brzeska, H., and Venyaminov, S. (1982) Tryptic fragments of calmodulin. $Ca^{2+}$- and $Mg^{2+}$-induced conformational changes. *J. Biol. Chem.* **257,** 11,584–11,590.
58. Minowa, O. and Yagi, K. (1984) Calcium binding to tryptic fragments of calmodulin. *J. Biochem. (Tokyo)* **96,** 1175–1182.
59. Dalgarno, D. C., Klevit, R. E., Levine, B. A., Williams, R. J., Dobrowolski, Z., and Drabikowski, W. (1984) 1H NMR studies of calmodulin. Resonance assignments by use of tryptic fragments. *Eur. J. Biochem.* **138,** 281–289.
60. Aulabaugh, A., Niemczura, W. P., and Gibbons, W. A. (1984) High field proton NMR studies of tryptic fragments of calmodulin: a comparison with the native protein. *Biochem. Biophys. Res. Commun.* **118,** 225–232.
61. Maune, J. F., Klee, C. B., and Beckingham, K. (1992) $Ca^{2+}$ binding and conformational change in two series of point mutations to the individual $Ca^{(2+)}$-binding sites of calmodulin. *J. Biol. Chem.* **267,** 5286–5295.

62. Beckingham, K. (1991) Use of site-directed mutations in the individual $Ca^{2(+)}$-binding sites of calmodulin to examine $Ca^{2(+)}$-induced conformational changes. *J. Biol. Chem.* **266,** 6027–6030.
63. Kilhoffer, M. C., Roberts, D. M., Adibi, A. O., Watterson, D. M., and Haiech, J. (1988) Investigation of the mechanism of calcium binding to calmodulin. Use of an isofunctional mutant with a tryptophan introduced by site-directed mutagenesis. *J. Biol. Chem.* **263,** 17,023–17,029.
64. Kilhoffer, M. C., Kubina, M., Travers, F., and Haiech, J. (1992) Use of engineered proteins with internal tryptophan reporter groups and pertubation techniques to probe the mechanism of ligand-protein interactions: investigation of the mechanism of calcium binding to calmodulin. *Biochemistry* **31,** 8098–8106.
65. Pedigo, S. and Shea, M. A. (1995) Discontinuous equilibrium titrations of cooperative calcium binding to calmodulin monitored by 1-D 1H-nuclear magnetic resonance spectroscopy. *Biochemistry* **34,** 10,676–10,689.
66. Pedigo, S. and Shea, M. A. (1995) Quantitative endoproteinase GluC footprinting of cooperative Ca2+ binding to calmodulin: proteolytic susceptibility of E31 and E87 indicates interdomain interactions. *Biochemistry* **34,** 1179–1196.
67. Shea, M. A., Verhoeven, A. S., and Pedigo, S. (1996) Calcium-induced interactions of calmodulin domains revealed by quantitative thrombin footprinting of Arg37 and Arg106. *Biochemistry* **35,** 2943–2957.
68. Sorensen, B. R. and Shea, M. A. (1998) Interactions between domains of apo calmodulin alter calcium binding and stability. *Biochemistry* **37,** 4244–4253.
69. Kilhoffer, M. C., Roberts, D. M., Adibi, A., Watterson, D. M., and Haiech, J. (1989) Fluorescence characterization of VU-9 calmodulin, an engineered calmodulin with one tryptophan in calcium binding domain III. *Biochemistry* **28,** 6086–6092.
70. Haiech, J., Kilhoffer, M. C., Craig, T. A., Lukas, T. J., Wilson, E., Guerra-Santos, L., and Watterson, D. M. (1990) Mutant analysis approaches to understanding calcium signal transduction through calmodulin and calmodulin regulated enzymes. *Adv. Exp. Med. Biol.* **269,** 43–56.
71. Haiech, J., Derancourt, J., Pechere, J. F., and Demaille, J. G. (1979) Magnesium and calcium binding to parvalbumins: evidence for differences between parvalbumins and an explanation of their relaxing function. *Biochemistry* **18,** 2752–2758.
72. Lafitte, D., Capony, J. P., Grassy, G., Haiech, J., and Calas, B. (1995) Analysis of the ion binding sites of calmodulin by electrospray ionization mass spectrometry. *Biochemistry* **34,** 13,825–13,832.
73. Heizmann, C. W. and Cox, J. A. (1998) New perspectives on S100 proteins: a multi-functional $Ca^{(2+)}$-, $Zn^{(2+)}$- and $Cu^{(2+)}$-binding protein family. *Biometals* **11,** 383–397.
74. Declercq, J. P., Tinant, B., Parello, J., and Rambaud, J. (1991) Ionic interactions with parvalbumins. Crystal structure determination of pike 4. 10 parvalbumin in four different ionic environments. *J. Mol. Biol.* **220,** 1017–1039.
75. Gilli, R., Lafitte, D., Lopez, C., Kilhoffer, M., Makarov, A., Briand, C., and Haiech, J. (1998) Thermodynamic analysis of calcium and magnesium binding to calmodulin. *Biochemistry* **37,** 5450–5456.

# 4

# Absorption and Circular Dichroism Spectroscopy

**Stephen R. Martin and Peter M. Bayley**

## 1. Introduction

Only four intrinsic protein chromophores absorb light significantly in the near-UV region of the spectrum (340–255 nm): the side chains of Trp, Tyr, Phe, and cystine (note: cysteine residues make no significant contribution). The absorption spectra are shown in **Fig. 1**. Although several amino acid side chains (notably Tyr, Trp, Phe, His, and Met) absorb light strongly in the far-UV region (below 250 nm), the most important contributor here is the peptide bond (amide chromophore), with $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions at approx 220 nm and approx 190 nm, respectively. The contribution of any individual chromophore to the total absorbance of the protein will depend, to some extent at least, upon its environment. The experimentally measured parameter, the absorbance A is related to the molar extinction coefficient, $\varepsilon_M$ ($M^{-1}\text{cm}^{-1}$), the path length $l$ (cm), and the protein concentration $C$ ($M$) by the Beer-Lambert law $A = \varepsilon_M.C.l$.

In circular dichroism (CD), the experimentally measured parameter is the difference in absorbance for left and right circularly polarized light, $\Delta A$ ( $= A_L - A_R$). Because CD is also an absorption phenomenon, the chromophores that contribute to the CD spectrum are the same as those contributing to a conventional absorption spectrum. The near-UV CD bands of proteins (deriving from Trp, Tyr, Phe, and cystine) reflect the tertiary and quaternary structure of the protein. The far-UV CD bands (deriving principally from peptide bond absorption) reflect the secondary structure of the protein (α-helix, β-sheet, β-turn, and random or unordered) (*see* **Subheading 3.5.**). The molar CD extinction coefficient, $\Delta\varepsilon_M$ ($= \varepsilon_L - \varepsilon_R$: units $M^{-1}\text{cm}^{-1}$) is calculated from the CD version of the Beer-Lambert law: $\Delta A = \Delta\varepsilon_M.C.l$.

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

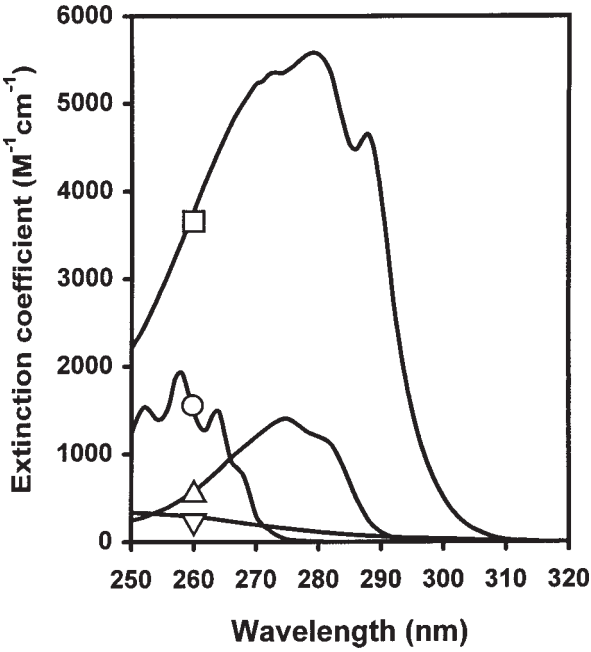


Fig. 1. Absorption spectra of tryptophan (□), tyrosine (△), phenylalanine (○), and cystine (▽) recorded in 10 m*M* phosphate buffer (pH 7.0). The spectrum for phenylalanine has been multiplied by 10 for clarity.

CD spectroscopy is widely used in the study of proteins because CD spectra are remarkably sensitive to molecular conformation. Although CD provides only low-resolution structural information it does have two great strengths. First, it is extremely sensitive to *changes* in conformation, whatever their origin, and second, a very wide range of solvent conditions is accessible to study with very small amounts of material. The principal applications of CD spectroscopy in the study of proteins are:

1. In the estimation of protein secondary structure content.
2. In detecting conformational changes brought about by changes in pH, salt, and added cosolvents (simple alcohols, tri-fluoroethanol, and so on).
3. In monitoring protein denaturation brought about by changes in temperature or by the addition of chemical denaturants (urea, guanidine hydrochloride).
4. In monitoring protein–ligand, protein–peptide, and protein–protein interactions.
5. In studying protein self-association through CD studies as a function of concentration.
6. In studying (in favorable cases) the kinetics of ligand binding (particularly slow dissociation processes), protein denaturation, and protein refolding.

There are numerous reviews that describe the principles of CD spectroscopy and its applications in the study of different biomolecules ***(1–5)***.

## 2. Materials

1. A CD instrument. The principal suppliers are: Jasco Inc. (Easton, MD, or Jasco U.K. Limited, Great Dunmow, Essex, U.K.); Jobin-Yvon (Longjumeau, France or Instruments S.A. U.K. Ltd., Stanmore, Middlesex, U.K.); Aviv and Associates (Lakewood, NJ); and On-Line Instrument Systems Inc. (Bogart, GA).
2. A set of quartz cuvets (either rectangular or cylindrical) with path lengths ranging from 0.1 to 10 mm. Self-masking (black-walled) micro or semimicro cuvets with 10 mm path length are particularly useful for near-UV CD and absorption measurements with small volumes (approx 0.25 mL). Cuvets are obtainable from several suppliers (e.g., HELLMA).
3. A sample of d-10-camphorsulfonic acid (d10-CSA, Aldrich) for instrument calibration.
4. All other standard reagents should be of the highest purity available. Organic solvents should be of spectroscopic grade (e.g., SpectrosoL from Merck) and should be checked for the absence of absorbing impurities.

## 3. Methods

### 3.1. Preparation of Instrument and Care of Cuvets

1. The instrument should always be purged with high-purity oxygen-free nitrogen (approx 3–5 L/min) for at least 20 min before starting the light source and while making measurements. Any oxygen present may be converted to ozone by far-UV light from the high intensity arc, and ozone will damage optical surfaces. Higher nitrogen flow rates should be used for measurements below 190 nm.
2. The instrument should be regularly calibrated. Prepare a solution of d10-CSA in water (approx 2.5 m*M*) and determine the precise concentration (*C*) by absorption spectroscopy using $\varepsilon_{285} = 34.5\ M^{-1}cm^{-1}$ (do not calculate the concentration by weight because the solid is hygroscopic). The calculated intensity at 290.5 nm in a 10-mm path length cuvet (in millidegrees) = $32980.C.\Delta\varepsilon_{M,290.5}$, where $\Delta\varepsilon_{M,290.5} = 2.36\ M^{-1}cm^{-1}$ ***(6)***. If the intensity is not within 1% of the expected value refer to the manufacturers's handbook for details of the adjustment procedure. It is also advisable to check the wavelength calibration of the instrument and its general far-UV transmission performance from time to time (*see* **Note 1**).
3. Cuvets with path lengths of 1 mm or less should always be calibrated. This is easily done using any solution with accurately known absorbance. Cuvets may have some strain that gives significant CD artifacts. Moderate strain can be tolerated, but it is sensible to eliminate any strain effects by always orienting the cuvet the same way in the CD instrument.
4. Cuvets should always be cleaned immediately after use, using a preparation such as HELLMANEX II cuvet cleaning solution (HELLMA). After cleaning, rinse extensively with distilled water, then ethanol, and dry using an air pump or by

evaporation. Cuvets should be stored in the cases generally provided by the manufactures.

### 3.2. Determination of Sample Concentration

1. Accurate sample concentrations are absolutely essential for the analysis of far-UV CD spectra for secondary structure content and whenever one wishes to make meaningful comparisons between different protein samples. Lowry or Bradford analyses are not sufficiently accurate for use with CD measurements unless they have been carefully calibrated using concentrations determined using a more direct method, such as quantitative amino acid analysis of the protein under investigation. We routinely determine protein concentrations using absorption spectroscopy as described in the **Subheading 3.2, step 2**.
2. Record the instrument baseline (450–250 nm) using a buffer solution that is *exactly* the same as that in which the protein is dissolved (*see* **Note 2**). Clean and dry the cuvet and record the spectrum of the sample with baseline subtraction and with temperature control. If the spectrum shows significant light scattering, i.e., significant background absorption above approx 315 nm, a correction should be applied. In most cases it is reasonable to assume that the scattering is Rayleigh in nature and that the absorbance caused by scatter is proportional to $\lambda^n$ (where the exponent $n$ is generally close to 4). The light-scattering contribution to be subtracted at 280 nm, for example, would then be $(A_{350nm})(350/280)^4 = (A_{350nm})(2.442)$ (*see* **Note 3**). When the extinction coefficient is known (*see* **Subheading 3.2, step 3**) the concentration can be calculated with considerable accuracy. Highly scattering samples should always be clarified by low-speed centrifugation or filtration prior to concentration determination.
3. Although it is possible to calculate the extinction coefficient of a protein with reasonable accuracy (***7,8***; *see* **Note 4**) it is much more reliable to measure it. This is best done using the Edelhoch method ***(8)***. Make identical dilutions of the protein stock in the experimental buffer and in the same buffer containing 6 *M* guanidine hydrochloride and record absorption spectra with appropriate buffer subtraction. Correct for light scattering, if necessary (*see* **Subheading 3.2, step 2**), and measure the absorbance at the chosen wavelength. Then, for example, the extinction coefficient at 280 nm is calculated from the amino acid composition as ***(8)***:

$$\varepsilon_{280,buffer} = (A_{280,buffer})(\varepsilon_{280,GuHCl})/(A_{280,GuHCl})$$

where $\varepsilon_{280,\,GuHCl}$ $(M^{-1}cm^{-1})$ = (#Trp)(5685) + (#Tyr)(1285) + (#cystine)(125).

In the case of calcium-binding proteins it is, of course, useful to perform this measurement for both the calcium-free and calcium-saturated forms. Also, one should not assume that the extinction coefficient of a protein is independent of temperature. For example, the extinction coefficient of apocalmodulin decreases by 5% on heating from 15 to 30°C, owing to instability of the C-terminal domain.

### 3.3. Sample Preparation

1. Samples should, of course, be of he highest possible purity. Near-UV CD signals, in particular, can be seriously distorted by the presence of relatively small amounts of protein impurities (if they have intense signals) and by the presence of nucleic acids, which have intense CD bands in this region.
2. Far-UV CD spectra of proteins (260–178 nm) are intense and small amounts of material are required to record them. Because all peptide bonds contribute to the spectrum the amount of material required is effectively the same for any protein. Typical quantities are 200 μL of a 0.1–0.15 mg/mL solution with a 1-mm path length cuvet or 30 μL of a 1.0–1.5 mg/mL solution with a 0.1 mm (demountable) cuvet. The latter is preferable for good far-UV penetration (*see* **Subheading 3.3 step 4**) but the material is not generally recoverable.
3. Near-UV CD spectra (340–255 nm) are much less intense than far-UV spectra and recording them requires more material. Spectra are usually recorded under conditions similar to those used for measuring a conventional absorption spectrum, e.g., use a 10-mm cuvet and aim for an absorbance at 280 nm in the range 0.7–1.0. Less-concentrated solutions may be used if the CD signals are intense.
4. CD signals will be seriously distorted if too little light reaches the photomultiplier. In practical terms, this means that one cannot make reliable measurements on samples with an absorbance (sample plus solvent) much greater than 1. The absorption spectrum of the sample should always be checked to see if (and where, *see* **Subheading 3.4 step 3**) this absorbance limit is exceeded. In far-UV measurements, the absorbance of the protein itself is generally rather small and the major problems arise from absorption by buffer components, almost all of which will limit far-UV penetration to some extent (*see* **Note 5**).

### 3.4. Data Collection

1. Set the scan speed and time constant. The product of the time constant and the scan speed should always be less than 0.5 nm. Higher values will give errors in both band position and band intensity (*see* **refs. *6*,*9–11*** for further discussion of errors in CD measurements). Typical parameters are a scan rate of 100 nm/min and a time constant of 0.25 s. Collecting multiple scans will improve the signal to noise (S/N) ratio to acceptable levels: the S/N ratio is proportional to the square root of the number of scans and to the square root of the time constant.
2. Set the spectral bandwidth. Increasing the spectral bandwidth reduces noise by increasing light throughput. The bandwidth should always be 2 nm or less to avoid distorting the spectrum. It may be necessary to use lower values in order to resolve fine structure in near-UV spectra.
3. Set the wavelength range. Far-UV spectra should generally be scanned from 260 to the lowest possible wavelength. This low-wavelength limit will depend largely upon the buffer being used (*see* **Subheading 3.3 step 4**). Near-UV spectra are routinely scanned over the range 340–255 nm.

4. Set the temperature control. Far-UV spectra in particular generally show some temperature dependence, even outside the range of any thermally induced unfolding of the protein.
5. Run a single scan to check that the selected parameters are appropriate. CD spectra will be seriously distorted if the photomultiplier voltage rises above a certain limit, generally of the order of 600 V. The low wavelength limit for far-UV spectra should be reset to a higher value if this photomultiplier voltage limit is exceeded. If the voltage is too high in the near-UV region either the protein concentration or the path length should be reduced. In the case of near-UV spectra, there is generally no CD signal in the 315–340 nm region. A significantly sloping baseline (often becoming increasingly negative toward lower wavelength) may indicate that there is a disulfide contribution to the spectrum. A sloping baseline may also be observed if the sample scatters light to a significant extent. Both factors should be checked by absorbance measurements (a CD signal outside the region of any absorption bands must be a result of scattering artifacts) and the upper wavelength limit for the scan should be extended (380 nm is generally sufficient).
6. Perform the measurement with signal averaging of a number of scans (5–10 is generally sufficient). If necessary, make any additions to the cuvet and repeat the measurement (*see* **Notes 6** and **7**).
7. Scan the baseline (with the same cuvet and buffer — *see* **Note 8**) using the same instrument settings. Do not be tempted to reduce the number of scans, because any noise in the baseline scan will simply be added to the sample scan in subsequent numerical processing.
8. For many purposes (e.g., in a titration or a denaturation experiment) it is often sufficient to have values of a CD signal at a single wavelength. To allow proper baseline alignment at higher wavelength, these values should be taken from full-wavelength scans whenever the signal is weak. When the signal is strong, the baseline alignment problems should be minimal and a single wavelength reading may be adequate. For far-UV titrations (e.g., with ligands or denaturants) and thermal unfolding experiments (*see* **Notes 9** and **10**), it may be helpful to use a solution at the normal concentration for a far-UV measurement (i.e., 0.1–0.15 mg/mL), but use a 10-mm path length cuvet. This restricts the accessible lower wavelength range, but normally permits measurements in the region of interest (generally 220 nm).

### 3.5. Data Analysis and Interpretation

1. Subtract the baseline scan from the sample scan. All spectra should have been collected with a starting wavelength that gives at least 15–20 nm at the start of the scan where the signal is zero (*see* **Subheading 3.4.**). After baseline subtraction this region should be (and usually is) flat but the signal may not be zero. This is usually caused by vertical drift in the signal. The solution is to average the apparent signal over the first 15–20 nm and subtract this average value from the whole curve.

2. Convert the spectrum to the desired units. The observed signal $S$ (in millidegrees) should generally be converted to the molar CD extinction coefficient ($\Delta\varepsilon_M$) or the mean residue CD extinction coefficient ($\Delta\varepsilon_{mrw}$) using:

$$\Delta\varepsilon_M = S/(32980.C_M.l) \text{ or } \Delta\varepsilon_{mrw} = S.mrw/(32980.C_{mg/mL}.l) \quad (\text{Units: } M^{-1}cm^{-1})$$

where $l$ is the path length (in cm), $C_M$ is the molar concentration, $C_{mg/mL}$ is the concentration in mg/mL, and mrw is the mean residue weight (molecular weight divided by the number of residues, *see* **Note 11**). Averaging far-UV intensities over the total number of amino acid residues in this way facilitates comparison between proteins. Averaging near-UV intensities in this way is not justified because only four different amino acid side chains contribute to the CD in this region. CD intensities are sometimes reported as molar ellipticity ($[\theta]_M$) or mean residue ellipticity ($[\theta]_{mrw}$), which may be directly calculated as

$$[\theta]_M = S/(10.C_M.l) \text{ or } [\theta]_{mrw} = S.mrw/(10.C_{mg/mL}.l) \quad (\text{Units: degrees.cm}^2\text{dmol}^{-1})$$

$[\theta]$ and $\Delta\varepsilon$ may be interconverted using the relationship $[\theta] = 3298.\Delta\varepsilon$

3. Near-UV CD bands from individual residues in a protein may be either positive or negative and may vary dramatically in intensity. Residues that are immobilized and/or interact strongly with neighboring aromatic residues produce the strongest signals. The near-UV CD spectrum of a protein does not allow one to say anything in detail about the tertiary structure of the protein. Knowledge of the position and intensity of CD bands expected for a particular residue is helpful in understanding the near-UV CD spectrum. The principal features are ***(12–14)***:
   - Phenylalanine has sharp fine structure in the range 255–270 nm with peaks generally observed at 262 and 268 nm ($\Delta\varepsilon_M \pm 0.3\ M^{-1}cm^{-1}$).
   - Tyrosine generally has a maximum in the range 275–282 ($\Delta\varepsilon_M \pm 2\ M^{-1}cm^{-1}$), possibly with a shoulder some 6 nm to the red.
   - Tryptophan often shows fine structure above 280 nm in the form of two $^1L_b$ bands (one at 288 to 293 and one some 7 nm to the blue, with the same sign $\Delta\varepsilon_M \pm 5\ M^{-1}cm^{-1}$) and a $^1L_a$ band (around 265 nm) with little fine structure ($\Delta\varepsilon_M \pm 2.5\ M^{-1}cm^{-1}$).
   - Cystine CD begins at long wavelength (>320 nm) and shows one or two broad peaks above 240 nm ($\Delta\varepsilon_M \pm 1\ M^{-1}cm^{-1}$), the long wavelength peak is frequently negative.

   Many of these features are illustrated in **Fig. 2**, which shows near-UV CD spectra of apo-calmodulin, $Ca_4$-calmodulin and the complex of the latter with an 18-residue peptide containing a single tryptophan residue. *Drosophila* calmodulin contains nine phenylalanines (which give the sharp bands at 262 and 268 nm) and a single tyrosine in the *C*-terminal domain (giving the broad band around 275 nm). These spectra show the profound change in the CD signal from Tyr-138, which may be used to monitor calcium binding to the *C*-terminal domain of calmodulin. The free peptide, which is unstructured, shows only a very weak CD signal, but immobilization in the complex generates peaks characteristic of tryptophan at 286

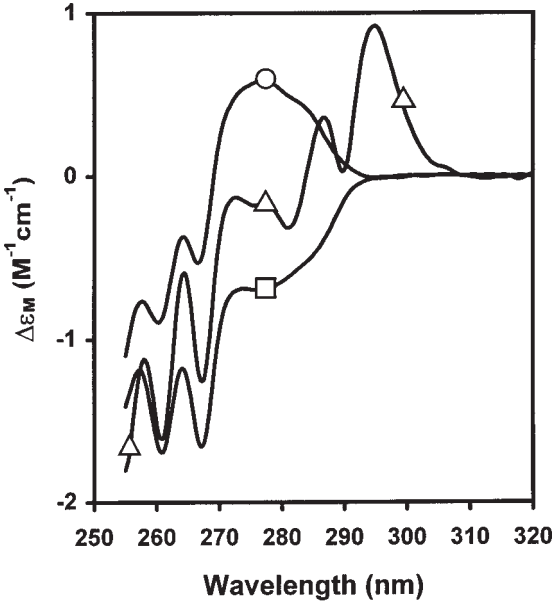


Fig. 2. Near-UV CD spectra of apo-calmodulin (○), $Ca_4$-calmodulin (□), and the complex of $Ca_4$-calmodulin with an 18-residue peptide from the calmodulin binding domain of skeletal myosin light-chain kinase (△).

and 293 nm. Studies with nontryptophan-containing peptides show that the spectrum of $Ca_4$-calmodulin remains effectively unchanged in the complex.

4. Far-UV CD spectra depend upon the secondary structure content of the protein and are generally easier to interpret. Characteristic features of the spectra of different protein classes may be summarized as follows ***(15)***:
   - All-α proteins show an intense negative band with two peaks (208 and 222 nm) and a strong positive band (191–193 nm). The intensities of these bands reflect α-helical content. $\Delta\varepsilon_{mrw}$ values for a totally helical protein would be of the order of –11 $M^{-1}cm^{-1}$ (208/222 nm) and +21 $M^{-1}cm^{-1}$ (191–193 nm).
   - Regular all-β proteins usually have a single negative band (210–225 nm, $\Delta\varepsilon_{mrw}$ –1 to –2.5 $M^{-1}cm^{-1}$) and a stronger positive band (190–200 nm, $\Delta\varepsilon_{mrw}$ 2–6 $M^{-1}cm^{-1}$). Intensities are significantly lower than for all-α proteins.
   - Unordered peptides and denatured proteins have a strong negative band (195–200 nm, $\Delta\varepsilon_{mrw}$ –4 to –8 $M^{-1}cm^{-1}$) and a much weaker band (either negative or positive) between 215 and 230 nm ($\Delta\varepsilon_{mrw}$ +0.5 to –2.5 $M^{-1}cm^{-1}$).
   - α+β and α/β proteins generally have spectra dominated by the α-helical component and, therefore, often show bands at 222, 208, and 190–195 nm. In some cases, there may be a single broad minimum between 210 and 220 nm because of overlapping α-helical and β-sheet contributions. Intensities depend on the α-helical content.

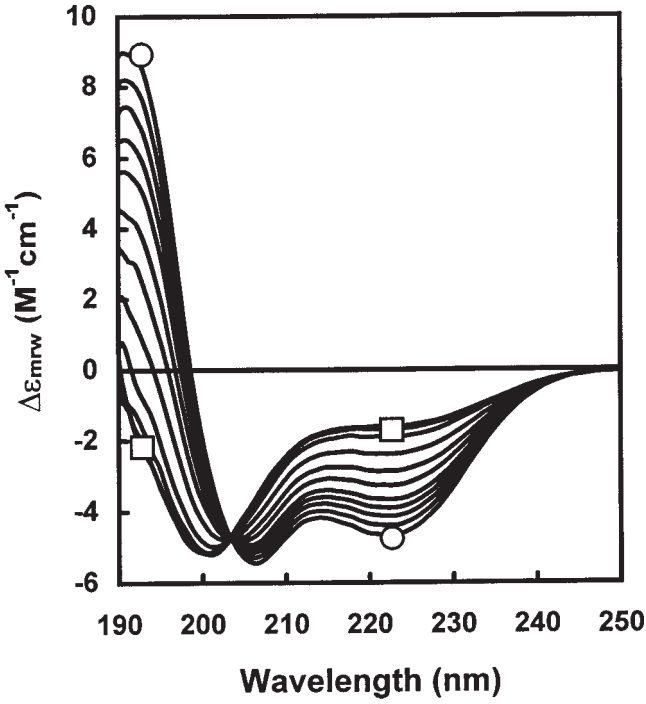


Fig. 3. Far-UV CD spectra of apo-calmodulin recorded at five-degree intervals over the temperature range 15 (○) to 75°C (□).

Some of these features are illustrated in **Fig. 3**, which shows the far-UV CD spectrum of apo-calmodulin as a function of temperature in the range 15–75°C. At low temperature the spectrum shows bands characteristic of the α-helix; heating causes progressive unfolding of the protein, the α-helical bands are lost and the bands characteristic of random or disordered structure appear.

5. Several approaches have been employed in attempts to determine the secondary structure content of proteins from their far-UV CD spectra (for reviews, *see* **refs. *1,2,4,5,15***). Early methods attempted to analyze CD spectra as linear combinations of reference (or basis) spectra for individual secondary structure elements that were derived from the spectra of model polypeptides or proteins. More modern methods analyze experimental CD curves more directly as linear combinations of the spectra of proteins whose structure has been determined by X-ray diffraction. All methods, even the oldest, give a reasonable estimate of α-helix content. CONTIN, VARSLC, and SELCON are all generally reliable; these, and other available methods have been discussed in several articles (***1,15–18***; *see* **Note 12**). The best starting point for anyone interested in these methods is the excellent review article by Greenfield ***(17)***. The validity of the various underlying assumptions in the calculation of secondary structure content from CD have been

discussed by Manning ***(19)***. One of the most important points to remember ***(17)*** is that, with the exception of nonconstrained least squares analysis, all the methods of analysis require a precise knowledge of protein concentration (*see* **Subheading 3.2.**).

6. When working with mutant proteins, it is essential to examine the effect of the mutation on the overall conformation and stability of the protein. CD provides a convenient means of doing this with limited amounts of material. Differences observed in the far-UV spectra are generally an indication that the mutation has produced a significant change in the secondary structure (*see* **Note 13**). However, differences observed in the near-UV region may derive from subtle changes in the environment of particular aromatic residues that are not necessarily associated with any major structural change.

## 4. Notes

1. d10-CSA has a second CD band at 192.5 nm ($\Delta\varepsilon_{M,192.5} = -4.72\ M^{-1}cm^{-1}$). The far-UV performance of a CD instrument can, therefore, be checked by recording the spectrum of d10-CSA (approx 5 m*M*) using a 1-mm path length cuvet. If the intensity ratio of the two peaks (–Signal[192.5]/Signal[290.5]) is significantly less than 1.95, then the machine is not performing correctly. This spectrum also provides a useful check on the wavelength calibration of the instrument.
2. This is particularly important if the solution contains any unusual components. For example, dithiothreitol (DTT) (oxidized), phenyl methyl sulfonyl fluoride (PMSF), and high concentrations of common chelators will distort the absorption spectrum if not accounted for.
3. A more elaborate method, easily implemented in a spreadsheet program, is to plot $\mathrm{Ln}(A_\lambda)$ against $\mathrm{Ln}(\lambda)$ and perform a least-squares fit to the straight line. This has the advantage that significant deviations from linearity may indicate the presence of contaminants rather than light scattering, and the actual value of the wavelength exponent ($n$) can be calculated.
4. Pace et al. ***(8)*** have shown that $\varepsilon_{280nm}$ can be predicted using

$$\varepsilon_{280nm}\,(M^{-1}cm^{-1}) = (\#Trp)(5500) + (\#Tyr)(1490) + (\#cystine)(125)$$

   This equation works best for proteins that contain tryptophan. For proteins that lack both *Trp* and Tyr one may use (with appropriate caution)

$$\varepsilon_{257.5nm}\,(M^{-1}cm^{-1}) = (\#Phe)(195) + (\#cystine)(295)$$

5. The majority of simple buffer components will permit far-UV CD measurements to below 200 nm. However, high concentrations of chloride and (especially) nitrate (use perchlorate if possible), certain solvents (dioxane, dimethyl sulfoxide [DMSO]), high concentrations (>25 m*M*) of some biological buffers (HEPES, PIPES, Mes), and high concentrations (>1 m*M*) of chelators (ethylene glycol-*bis* *N*,*N*,*N*',*N*'-tetraacetic acid [EGTA]/ethylenediaminetetracetic acid [EDTA]) should be avoided. It is also worth noting that distilled water stored in a polyeth-

ylene bottle will develop poor far-UV transparency owing to the presence of eluted polymer additives.

6. Making additions (especially to short path length cuvets) poses several problems. The small volume sample has to be mixed thoroughly by inversion or using a long thin pipet tip. This should be done with care in order to minimize the (almost inevitable) small loss of solution. Note that, with dilute samples, especially of small highly charged molecules, one can get loss of sample through absorption to pipet tips, and so on. It is generally wise to use the same pipet tip for all mixing operations in a single experiment. Finally, because additions may increase the absorption, it is always worth estimating (or better measuring) what the final absorbance will be.
7. When working with calcium-free proteins, make sure that the component being added is itself calcium free. If this is not possible by pretreating the solution with Chelex (for example, when using solutions of denaturants) then include some EGTA or EDTA in the solution. Remember that the interaction of these reagents with calcium is strongly pH-dependent. The $K_d$ values for both chelators are approx 10 n*M* at pH 7.8, but rise quickly above 1 μ*M* below pH 6.8 (EGTA) or pH 6.1 (EDTA).
8. Strictly speaking, the true baseline in CD should be the cuvet plus solvent with a sample that has the same normal absorption, but no CD. However, this is seldom done and is unlikely to be a problem except with very weak signals.
9. In thermal unfolding experiments, the temperature should be increased slowly (no more than 1°/min). The temperature should be measured using an immersible electronic probe in the cuvet rather than in the water bath, and reversibility on lowering the temperature should be checked. Buffers with high thermal coefficients (e.g., Tris-HCl) should be avoided if possible.
10. CD spectra can be recorded at temperatures below zero by using suitable water-alcohol or water-glycerol mixtures. It is essential in such studies to check for any direct effect of the solvent itself on the conformation of the protein. This is done by measuring the CD spectrum in the solvent at room temperature and comparing it with the spectrum measured in an aqueous buffer.
11. Large globular proteins generally have a mean residue weight of approx 111. The actual value should always be calculated.
12. The articles by Greenfield ***(17)*** and Venyaminov and Yang ***(15)*** provide useful lists of computer programs available for the determination of secondary structure content from CD. All of these are easily implemented on a PC.
13. A difference in the shape of the far-UV spectra of the wild-type and mutant protein almost certainly indicates a difference in conformation. However, if the spectra can be made identical by a simple multiplication then the difference very probably arises from small differences in concentration. A weaker signal for a mutant protein may indicate that the mutation has affected the stability of the protein and that the mutant is partially unfolded. Thermal or chemical denaturation experiments can be used to check for this possibility.

## References


1. Yang, J. T., Wu, C.-S. C., and Martinez, H. M. (1986) Calculation of protein conformation from circular dichroism. *Methods Enzymol.* **130,** 208–269.
2. Woody, R. W. (1995) Circular dichroism. *Methods Enzymol.* **246,** 34–71.
3. Woody, R. W. (1996) Theory of circular dichroism of proteins, in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., ed.), Plenum, New York, pp. 25–67.
4. Johnson, W. C., Jr. (1985) Circular dichroism and its empirical application to biopolymers. *Methods Biochem. Anal.* **31,** 61–163.
5. Johnson, W. C., Jr. (1988) Secondary structure of proteins through circular dichroism spectroscopy. *Annu. Rev. Biophys. Biochem.* **17,** 145–166.
6. Johnson, W. C., Jr. (1990) Protein secondary structure and circular dichroism: a practical guide. *Prot. Struct. Funct. Genet.* **7,** 205–214.
7. Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319–326.
8. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423.
9. Johnson, W. C., Jr. (1996) Circular dichroism instrumentation, in Circular dichroism and the conformational analysis of biomolecules (Fasman, G. D., ed.), Plenum, New York, pp. 635–652.
10. Hennessey, J. P., Jr. and Johnson, W. C., Jr. (1982) Experimental errors and their effect on analyzing circular dichroism spectra of proteins. *Anal. Biochem.* **125,** 177–188.
11. Martin, S. R. (1996) Circular dichroism, in *Proteins Labfax* (Price, N. C., ed.) BIOS Scientific Publishers Ltd., Oxford, pp. 195–204.
12. Strickland, E. H. (1974) Aromatic contributions to circular dichroism spectra of proteins. *CRC Crit. Rev. Biochem.* **2,** 113–175.
13. Woody, R. W. and Dunker, A. K. (1996) Aromatic and cystine side-chain circular dichroism in proteins, in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., ed.), Plenum, New York, pp. 109–157.
14. Woody, R. W. (1985) Circular dichroism of peptides, in *The Peptides*, vol. 7 (Hruby, V. J., ed.), Academic, New York, pp. 15–114.
15. Venyaminov, S. Y. and Yang, J. T. (1996) Determination of protein secondary structure, in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., ed.), Plenum, New York, pp. 69–107.
16. van Stokkum, I. H. M., Spoelder, H, J. W., Bloemendal, M., van Grondelle, R., and Groen, F. C. A. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem.* **191,** 110–118.
17. Greenfield, N. J. (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* **235,** 1–10.
18. Sreerama, N. and Woody, R. W. (1994) Protein secondary structure from circular dichroism spectroscopy. Combining variable selection principle and cluster analy-

sis with neural network, ridge regression and self-consistent methods. *J. Mol. Biol.* **242,** 497–507.

19. Manning, M. C. (1989) Underlying assumptions in the estimation of secondary structure content in proteins by circular dichroism spectroscopy — a critical review. *J. Pharm. Biomed. Anal.* **7,** 1103–1119.

# 5

# Fourier Transform Infrared Spectroscopy of Calcium-Binding Proteins

**Heinz Fabian and Hans J. Vogel**

## 1. Introduction

Infrared spectroscopy measures absorptions of vibrating molecules and yields information about molecular structure and structural interactions. Over the last two decades, the infrared technique has emerged as a very useful tool for examining protein conformation as a result of the increase in energy throughput, achievable signal-to-noise ratio, wavenumber accuracy, and data aquisition rates that came with the development of Fourier transform infrared (FTIR) spectrometers. High-quality infrared spectra can now rapidly be aquired and require only relatively small amounts of protein. The size of the protein or the nature of the environment does not limit the application of FTIR spectroscopy. Importantly, measurements of proteins in aqueous solution are almost routine now. Furthermore, the process of obtaining structural information is not restricted to a static picture, but can also be achieved in real time by applying time-resolved infrared techniques. The effects of environmental factors, point mutations, or ligand binding on the structure of the proteins can be examined with high sensitivity by using peptide backbone and side-chain infrared bands as conformation-sensitive monitors. In combination with isotope labeling, the technique also permits the study of protein–protein or protein–peptide interactions.

## 2. Materials

### *2.1. FTIR Spectrometer*

Spectrometers required for measuring high-quality spectra of proteins in the midinfrared region are available from a variety of manufacturers. These research-grade instruments offer a spectral resolution of better than 1 $cm^{-1}$ and

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

are equipped with sensitive infrared detectors. The most common detectors are triglycine sulfate (TGS), deuterated triglycine sulfate (DTGS), and liquid-nitrogen cooled mercury-cadmium telluride (MCT). MCT detectors are more sensitive and permit much higher aquistion rates than TGS and DTGS detectors. Nevertheless, MCT detectors are not always the best choice because they suffer from problems associated with detector nonlinearity at high absorbance values. The limited linear range may be a disadvantage when studying proteins dissolved in water because of the high absorptivity of water bands in certain regions. TGS and DTGS detectors work at room temperature and have the advantage of a significantly extended detector linearity. Although the latter detectors require longer aquisition times, they are a good choice for many experiments under equilibrium conditions. Some research-grade FTIR spectrometers are able to accomodate two detectors and allow for rapid computer-controlled exchange between these detectors (*see* **Note 1**).

## 2.2. Sampling Devices and Sample Handling

The majority of FTIR experiments with calcium-binding proteins have been carried out in aqueous solution and were conducted with conventional transmission geometries. Here, the nature of the material of which the cell window is constructed and the pathlength of the cell are important.

### 2.2.1. Window Materials

Among the IR window materials available for experiments in aqueous solution, calcium fluoride ($CaF_2$) is the most common because (1) it has a low refractive index, which is similar to that of water; (2) it is relatively rugged; and (3) is transparent from the midinfrared (>1000 $cm^{-1}$) to the ultraviolet (UV) region of the spectrum. Barium fluoride ($BaF_2$), a similar window material, has a lower spectral cutoff than $CaF_2$ (800 $cm^{-1}$), but it is significantly more soluble in aqueous solution. Although $CaF_2$ is the most suitable window material for protein measurements, it is not an ideal window material for long-term measurements of $Ca^{2+}$-binding proteins. The solubility of $CaF_2$ in water is very low, but a possible contamination of the sample by dissolution of $Ca^{2+}$ from the $CaF_2$ window cannot be excluded. Consequenctly, the $Ca^{2+}$-free form of a sample may not persist during the collection of spectra over longer time scales. Window materials that are insoluble in water (such as KRS-5, ZnSe, or Irtran) are available, but are characterized by an unfortunate high refractive index, which results in major reflection losses and persistent fringing in the spectra (*see* **Note 2**).

### 2.2.2. Path Length of the Cell

The choice of the path length of the IR cell depends upon which region of the spectrum is of interest. The most useful probe of protein secondary struc-

ture is the amide I backbone mode, which occurs in the region 1610–1700 $cm^{-1}$. Studies of proteins in the amide I region are complicated by the fact that the bending vibration of water absorbs very strongly near 1640 $cm^{-1}$. As a consequence, for transmission measurements in the amide I region very short path length cells of 3–8 μm are needed. Experimentally, it is simpler to obtain protein spectra in deuterium oxide ($D_2O$) solution than in $H_2O$ solution. The infrared bands of $D_2O$ compared to those of $H_2O$ occur at lower wavenumbers because of the downshifted vibrations of the heavier deuterium atoms. This isotopic effect creates a region of relatively low absorbance between 1400 and 1800 $cm^{-1}$, a window for observing the weak infrared bands of the dissolved protein. Much longer path lengths of 40–80 μm may then be used (*see* **Note 3**).

### 2.2.3. Design of the IR Cells

Flowthrough demountable cells with luer-lock fittings and spacers covering path lengths of 6 to 200 μm are often used for protein FTIR measurements. These cells are available from virtually any infrared accessories supplier, but they have some disadvantages in practice such as being difficult to clean and giving rise to accidental injection of gas bubbles. To circumvent these problems, we use custom-made IR cells of different design that consist of a flat cover disk (typically made of $CaF_2$) and a second disk of the same material (sample disk), with the center deepened to form a recessed parallel surface surrounded by a trough (*see* **Fig. 1**). The trough prevents direct contact of the sample with the outer part of the disk. Pressing the cover disk onto the sample disk seals the cell, and this is sufficient to prevent the evaporation of water for many hours at room temperature. These windows are fitted into a metal jacket through which heating or cooling liquid from an external bath can circulate. For measurements at high temperatures and/or long-time experiments, the sealing surface of the disks is lubricated with mineral oil prior to filling and assembling of the cell. Depending upon the diameter and the depth of the recessed surface of the window (i.e., the path length of the cell), only a few microliters are required to fill the cell. Moreover, this type of cell is easy to fill with a solution, and can be assembled and disassembled (most of the solution can be recovered), and cleaned between measurements. It provides a constant path length, which is very difficult to achieve with conventional tin or teflon spacers.

### 2.2.4. Attenuated Total Reflection (ATR) Sampling Technique

For ATR measurements, the sample is prepared on the surface of an infrared transparent crystal. The IR beam is guided through the crystal in such a way that some total reflections take place at the surface. Because the IR beam penetrates slightly into the surrounding medium, the deposition of an infrared absorber on the crystal surface causes the infrared light to be partially absorbed.

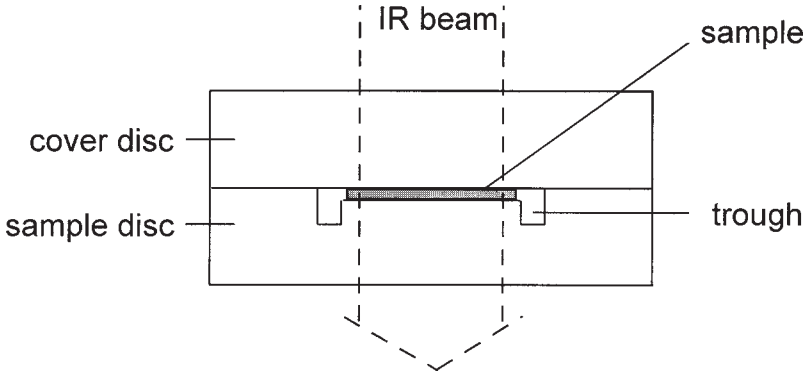


Fig. 1. Illustration of our custom-made IR cell.

The penetration depth of the infrared radiation in this arrangement is strictly dependent on the wavelength and, therefore, the infrared spectrum measured contains only information on a very thin layer of the sample that is in close proximity to the surface of the crystal. This allows one to obtain a spectrum of a protein in $H_2O$ solution without much interference from infrared absorption of the bulk water ***(1)***. Surface adsorption, however, may significantly change the secondary structure of the protein molecules, which are in direct contact with the crystal. Although the contribution of those moleclues to the total absorbance measured might by small, one should proceed with caution in structural studies of proteins by ATR spectroscopy (*see* **Note 4**).

### *2.2.5. Buffers and Denaturants*

Many commonly employed buffers, such as phosphate, cacodylate, Tris-HCl, or HEPES, are acceptable. Buffers containing carboxylic acid groups, such as acetate or carbonate buffers, are not ideal because their infrared absorption bands overlap with those for the protein backbone.

Measurements of proteins in the presence of the calcium chelators EDTA or EGTA are complicated by the fact that the carboxylic groups of these chelators have strong infrared bands in the region 1570–1630 $cm^{-1}$. In addition, the spectral characteristics of these bands are influenced by $Ca^{2+}$-binding. As a consequence, concentration and spectral features of the chelator in the sample cell and the reference cell must be perfectly matched in order to avoid a misinterpretation of the IR spectra between 1570–1630 $cm^{-1}$.

Obtaining infrared spectra of proteins in the presence of high concentrations of the most commonly used denaturating agents, urea and guanidinium chloride (GdmCl), is not simple. The strong infrared bands of urea around 1613 $cm^{-1}$ or GdmCl around 1600 $cm^{-1}$ mask the much weaker protein amide I band. Isoto-

pic labeling of the denaturant (e.g., use of $^{13}C$ = O-labeled urea) helps to circumvent this problem by shifting the urea band to 1560 $cm^{-1}$ ***(2)***.

## 3. Methods

A Fourier transform spectrometer does not directly measure the desired spectrum. The FTIR instrument is nondispersive and makes use of an interferometer to encode data from the whole spectral range simultaneously ***(3)***. A computer is required for controlling data collection, converting the signal-averaged interferogram to a single beam spectrum by means of Fourier transformation, and for subsequent computations that give the final absorbance spectrum that has been corrected for instrumental contributions. To obtain a high-quality spectrum of a protein, an accumulation of 100–200 interferograms recorded with a moderate resolution of 2 or 4 $cm^{-1}$ is often sufficient. FTIR spectroscopy is a single-beam technique and thus the protein and the solvent/buffer spectrum have to be measured separately.

### *3.1. Instrument Purge*

The aquisition of high-quality infrared spectra requires to reduce drastically the contributions of water vapor and $CO_2$ in the sample compartment of the spectrometer by purging the instrument with dry air or with nitrogen. Central pressure air available in large research institutions often does not fulfill the high-quality criteria necessary for FTIR. Here, the use of an extra air dryer in front of the spectrometer is strongly recommended.

### *3.2. Sample Preparation*

In case of soluble proteins, the purified and dry protein is weighed and then dissolved at the desired concentration in the buffer of choice (*see* **Note 5**). For measurements in $H_2O$ solution, relatively high protein concentrations (>10 mg/mL) are required. Much lower protein concentrations (>1 mg/mL) are required to obtain high-quality spectra of proteins in $D_2O$ solution, because the latter measurements allow for the use of cells with much longer path length (3–8 μm for $H_2O$ vs 40–80 μm for $D_2O$).

### *3.2. Data Collection/Data Manipulation*

The measurement starts by recording the background of the spectrometer through an empty position of the sample compartment, then data of the sample and the buffer are collected under identical conditions (such as number of scans, resolution, and so on). To obtain the spectrum of the protein, digital subtraction of solvent/buffer absorptions from the spectrum of the protein is required. For an appropriate subtraction, the spectrum of the solvent/buffer should be recorded under practically identical physicochemical parameters (such as tem-

perature, ionic strength, pH) because slight variations will cause changes in the spectral features of the $H_2O$ or $D_2O$ bands, thereby preventing an ideal subtraction of the buffer contributions. For example, for measurements in water, the temperatures of the sample in $H_2O$-solution and the buffer should coincide within 0.1°C in order to avoid artifacts caused by temperature differences. The subtraction of water from a protein spectrum requires a reference water band that does not overlap with those of the sample. The weak combination band of $H_2O$ around 2126 $cm^{-1}$ may serve as a good approximation to interactively subtract the water features. The final water subtraction should be performed using a different spectral region with stronger $H_2O$ absorption, such as the one in the vicinity of approx 3645 $cm^{-1}$ ***(4)***.

### *3.2.1. Hydrogen–Deuterium Exchange, a Specific Feature of Protein Studies in $D_2O$*

The hydrogen–deuterium exchange of amide protons can be monitored by the disappearence of the band characteristic of N–H bending near 1545 $cm^{-1}$ (amide II) and the appearence of N–D absorption near 1455 $cm^{-1}$ (amide II'). The shift of the amide I band (predominantly C = O stretching vibration mode of the amide group) upon deuteration of the backbone hydrogens (labeled amide I' by convention) is only small (5–10 $cm^{-1}$). However, individual spectral components of the amide I band of a protein often reveal different exchange kinetics. This greatly assists the assignment of band components arising from different secondary structural classes. Despite this positive aspect, it can also complicate the interpretation of the amide I' region, if a protein cannot completely be exchanged**. Figure 2A** (solid line) shows the infrared spectrum of the apo-form of calmodulin after complete H–D exchange. The solvent spectrum (**Fig. 2A**, dotted line) was measured in an optimally matched second cell of slightly reduced path length, which takes into account the slightly lower $D_2O$ concentration in the protein sample measured**. Figure 2B** (solid line) shows the corresponding buffer-subtracted spectrum of apo-calmodulin, together with a spectrum of the apo-form measured 15 min after dissolving the protein in $D_2O$ (dotted line). Residual intensity at 3300 $cm^{-1}$ (amide A; N–H stretching of the peptide groups) indicates that a number of amide protons are not exchanged after short exposure of apo-calmodulin to $D_2O$. The amide A is the best indicator for nonexchanged N–H groups because of the lack of other protein absorptions in the range 3200–3400 $cm^{-1}$. The same information cannot easily be deduced from the residual intensity in the amide II region, because infrared bands arising from amino acid side-chain groups overlap with the remaining amide II band. For example, in the case of apo-calmodulin, the carboxyl groups of aspartate and glutamate residues absorb around 1575 $cm^{-1}$.

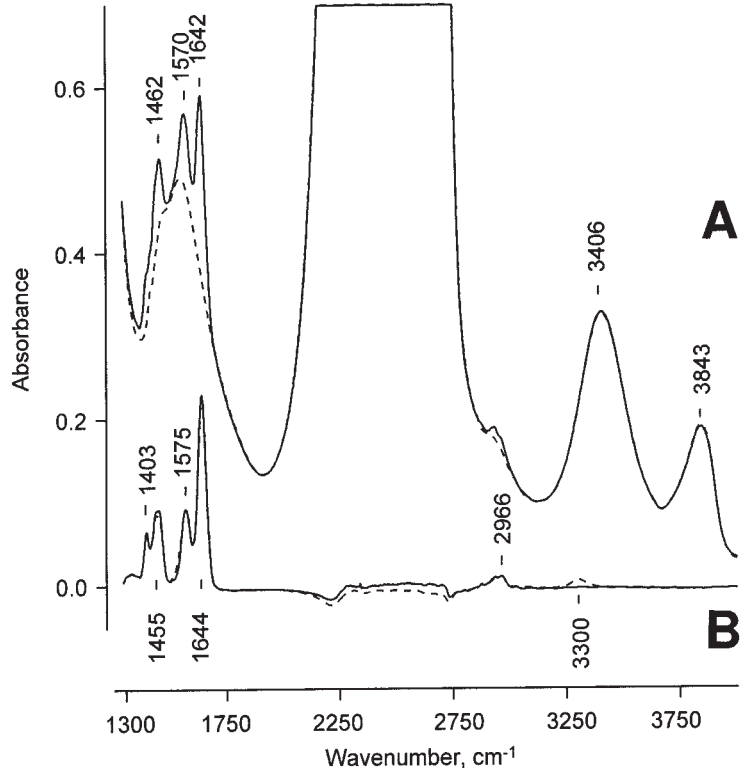


Fig. 2. **(A)** IR spectra of apo-calmodulin (*c* approx 11 mg/mL) in $D_2O$-buffer (100 m*M* Na-cacodylate, pH 7.0) after complete H/D exchange (*solid line*), together with the buffer spectrum (*dashed line*). For complete exchange of all amide protons with deuterons, the lyophylized protein was dissolved in $D_2O$-buffer and left overnight at room temperature. To eliminate the possibility of significant $Ca^{2+}$-leaking from the infrared cell, the apo protein sample was placed in the $CaF_2$ cell immediately before the measurement, which was completed within less than 10 min. Infrared spectra were recorded on a Bruker IFS-66 FTIR spectrometer equipped with a DTGS detector. For each sample, 128 interferograms were co-addded and Fourier-transformed applying a Happ-Genzel apodization function to generate a spectrum with a nominal resolution of 4 $cm^{-1}$. **(B)** IR spectrum of the fully exchanged apo-form after subtraction of the buffer spectrum (*solid line*). The dashed line shows a correponding difference spectrum of only partly exchanged apo-CaM recorded 15 min after dissolution of the lyophilized protein in $D_2O$.

### 3.2.2. Residual Water Vapor

It is almost impossible to remove all water vapor by purging of the spectrometer. In addition, the level always changes when the sample chamber is opened. It is therefore convenient to record spectra at low, but well-matched, levels of

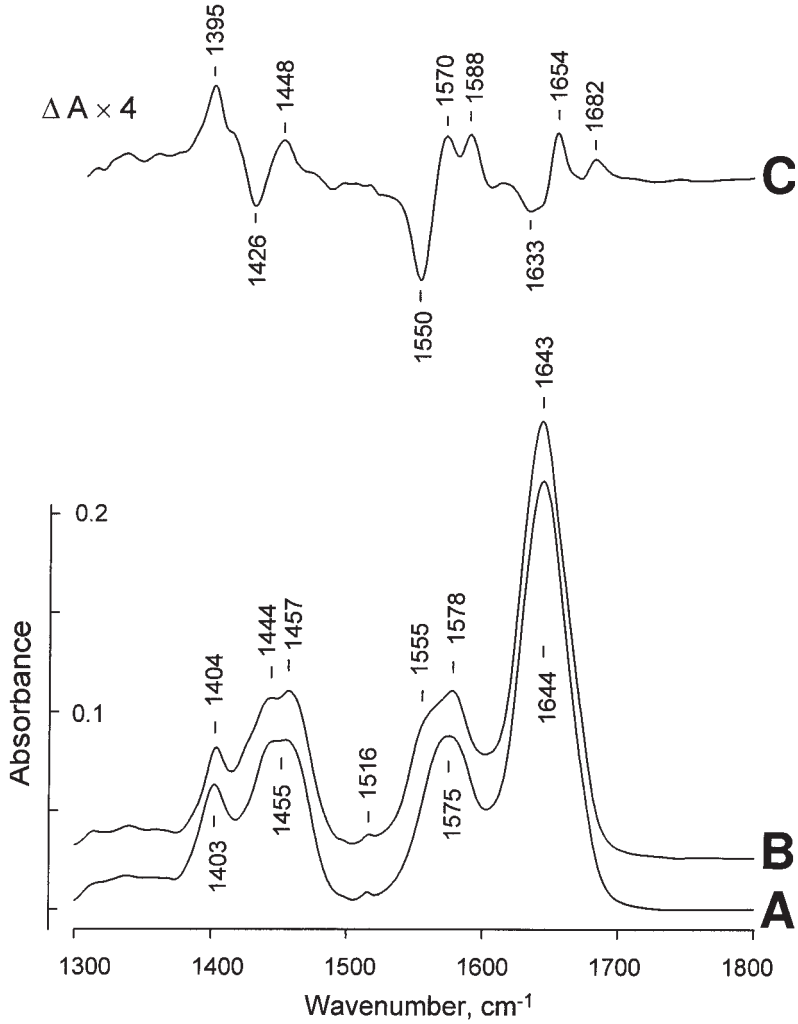


Fig. 3. IR spectrum of $Ca^{2+}$-free calmodulin **(A)** and of calmodulin saturated with $Ca^{2+}$ **(B)**, both spectra are shown after $D_2O$-buffer subtraction. **(C)** Infrared difference spectrum, obtained by subtracting the spectrum of the $Ca^{2+}$-bound form from the spectrum of the $Ca^{2+}$-free form. Note that the absorbance scale for the difference spectrum was expanded by a factor of 4.

water vapor for the sample and the reference. A sample shuttle that allows the background to be signal averaged almost concurrently with the sample is very helpful. Residual water vapor can then simply be subtracted from the sample spectrum by using prerecorded water vapor spectra. This is mandatory for pro-

tein measurements, because the narrow water vapor bands overlap with the conformation-sensitive amide I/II bands. For aqueous solutions, subtracting two buffer spectra from the same cell collected at different purge levels should generate the correct water vapor spectrum. Any over- or undersubtraction of water vapor can best be visualized by calculating the second derivatives of the spectra, which enhances narrow bands in particular ***(5)***. The subtraction factor must be varied until the second-derivative spectrum is featureless in the range 1750–1850 $cm^{-1}$, which is normally free of any protein bands.

## 3.3. Data Processing Techniques

### 3.3.1. IR Difference Spectroscopy

Difference spectroscopy involves the subtraction of a protein IR spectrum in state A from that of the protein in state B. The resultant difference IR spectrum only reveals features that are associated with those groups involved in a conformational change. **Figure 3** shows the infrared spectra of apo-calmodulin (trace A) and of calmodulin saturated with $Ca^{2+}$ (trace B). The two spectra are dominated by a strong band centered at 1643/1644 $cm^{-1}$, which is a result of the amide I' mode of calmodulin. The amide II' band is located at around 1455 $cm^{-1}$. Spectral features at 1550–1590 $cm^{-1}$ arises from the antisymmetric $COO^-$ stretching vibrations of the carboxylate moiety of the amino acid side-chain groups of glutamate and aspartate; the correponding symmetric bands are located at 1390–1430 $cm^{-1}$. The carboxylate modes are established markers of metal-ion binding (***6,7***; *see* also Chapter 13, Volume 1). Calcium binding results in an upshift of the symmetric stretching band and a downshift of the antisymmetric band. In calmodulin, 14 out of the 38 $COO^-$ groups are found in the $Ca^{2+}$ -binding sites, and the features at 1430–1390 and 1550–1590 $cm^{-1}$ in the infrared difference spectrum (*see* **Fig. 3C**) are highly characteristic of the spectral changes associated with calcium binding to the carboxylate ligands in calmodulin. Positive and negative features in the amide I' region suggest that only slight changes in secondary structure take place when $Ca^{2+}$ binds to calmodulin.

### 3.3.2. Derivation Methods and Fourier Deconvolution

A major problem in the use of the amide I mode to secondary structure analysis is that the infrared bands arising from each type of polypeptide conformation are inherently broad and lie close together. This leads to only weakly resolved broad features. Two mathematical procedures are very useful to help identify overlapping components within the composite amide I band contour. The first involves calculation of the *n*th derivative of the spectrum. Often, the second derivative is calculated, which gives a negative peak for every band or

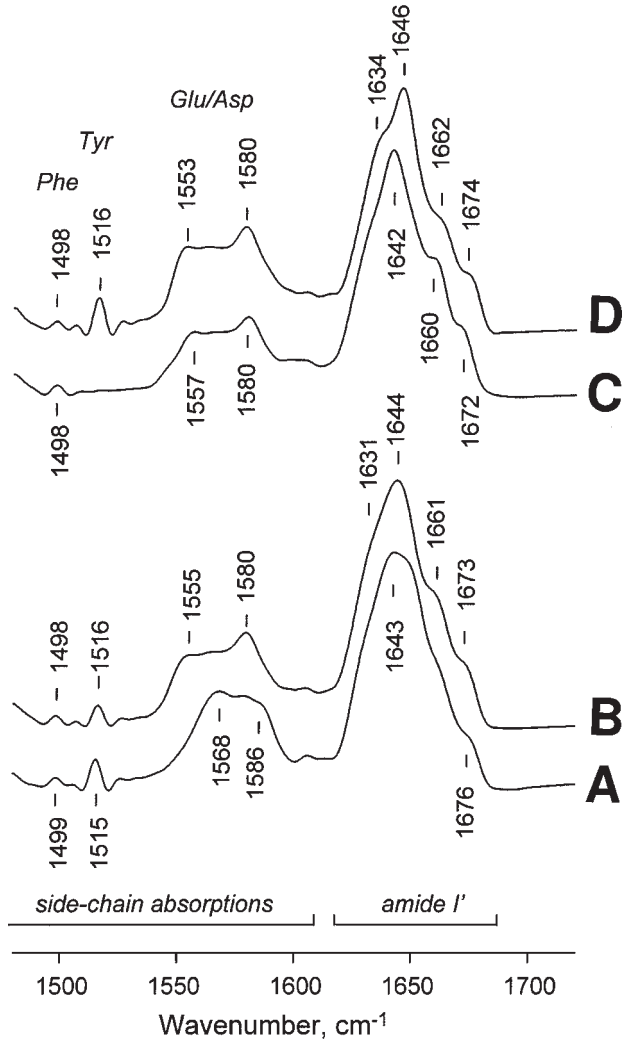


Fig. 4. Lower traces: **(A)** IR spectra of intact apo-calmodulin and **(B)** intact calmodulin saturated with $Ca^{2+}$. Upper traces: **(C)** IR spectra of the *N*-terminal domain (residues 1–77) and **(D)** the *C*-terminal domain (residues 78–148) of the $Ca^{2+}$-saturated form of calmodulin. All spectra are shown after band narrowing by Fourier self-deconvolution using a half bandwidth of 16 $cm^{-1}$ and a band-narrowing factor of 2.

shoulder in the spectrum. Because sharp bands are enhanced at the expense of broad ones, this process does not preserve the integreated areas of the individual components. The second band narrowing approach involves Fourier self-deconvolution ***(8)***, which also significantly improves the degree to which the

individual component bands can be resolved, whereby the relative integrated intensities are maintained. Both band-narrowing techniques greatly amplify features in the spectra originating from random noise and/or uncompensated water vapor; they need to be used with great care to avoid artifacts ***(5,9)***. A comparison of the deconvoluted spectra of the apo-form and the $Ca^{2+}$ form of calmodulin given in **Fig. 4A**,**B**, respectively, reveals several band components in the amide I' region, which are hidden in the original spectrum of the corresponding protein (*see* **Fig. 3A**,**B**). In addition, the visualization of the fine structure of the side-chain absorptions of glutamate and aspartate is improved. Moreover, weak bands at 1515/1516 and 1498/1499 $cm^{-1}$, which are caused by amino acid side-chain absorptions ***(10)*** of tyrosine and phenylalanine, respectively, can be identified in the deconvoluted spectra.

### 3.3.3. Curve Fitting of Band Contours

Curve fitting of amide I/I' band profiles is often used to quantitatively analyze underlying band components. In the curve-fitting approach, the number of component bands estimated by Fourier self-deconvolution and derivative spectra, plus their approximated width, height, and shape are used as input parameters in an iterative least squares procedure that attempts to reproduce the measured amide I/I' band profile by varying these parameters. For practical reasons, deconvoluted spectra should be subjected to curve fitting because least-square algorithms are significantly more reliable for spectra with an enhanced profile. When a reasonable fit is obtained, the fractional areas of the fitted components are taken as directly proportional to the relative quantities of the structure elements they represent. The percentages of different secondary structure elements can then be estimated by adding the areas of all component bands assigned to each of these structures and expressing the sum as a fraction of the total amide I/I' band area ***(11)***.

### 3.3.4. Problems Associated with the Curve-Fitting Procedure

The curve-fitting approach (like all curve-fitting applications) has some inherent problems. An element of subjectivity is the assumption that the number of band components estimated by self-deconvolution or derivation reflects the real number of components. In cases where bands significantly overlap, even the applied band-narrowing procedures certainly fail in separating the components present. Another assumption in this method is that the molar absorptivities of the bands associated with different secondary structural elements are identical, which is at best a rough approximation ***(9)***. A very critical step is the assignment of the component bands, which is based on theoretical calculations and on emperical spectra-structure correlations experimentally established for model polypeptides and proteins of known three-dimensional structure ***(9,11)***.

These correlations indicate that amide I' bands in the range 1650–1658 $cm^{-1}$ correspond to α-helices, and a component at approx 1645 $cm^{-1}$ to irregular parts of polypeptides backbones. Turns are associated with various component bands between 1660 and 1690 $cm^{-1}$. One or more bands between 1620 and 1635 $cm^{-1}$ can be attributed to β-sheet structures, the antiparallel type can be identified by the presence of another weaker band near 1675–1695 $cm^{-1}$. Some proteins, however, contain secondary structures that absorb outside these frequency ranges. Amide I' bands that result from α-helical structure may also be present below 1650 $cm^{-1}$, in some cases even near 1630–1640 $cm^{-1}$ with highly solvent-exposed helices ***(12)***. Calmodulin, parvalbumin and troponin C belong to those proteins known to be highly α-helical but exhibiting an amide I' band centered at approx 1645 $cm^{-1}$ ***(7,13,14)***. Without structural information provided by other techniques, the spectra of these proteins could easily be mistaken for predominantly irregular structures according to the standard assignment of amide I' bands given above. In addition, the amide I' bands of the amino- and carboxy-terminal domains of calmodulin ***(15)***, both known to have similar α-helical structures, are different (compare **Fig. 4C**,**D**). This renders the amide I' band for the intact protein very broad (*see* **Fig. 4B**) in comparison to that of other proteins that contain a high percentage of α-helix. Such a broad and less-structured amide I' bandshape creates a large degree of subjectivity in the quantitative estimation of protein secondary structure from infrared spectra using curve-fitting procedures.

### 3.3.5. Pattern-Recognition Approaches

Pattern-recognition methods are a quite different approach to estimate the secondary structure of a protein. These methods use infrared spectra of proteins with known 3D structure as a calibration matrix (for a review, *see* **ref. *1***) and are analogous to well known procedures used in the analysis of CD spectra. An advantage of the pattern recognition approaches is that they do not require the assignment of individual component bands to different types of secondary structure. This approach is, however, dependent on the reference database, which is still rather limited (primarily soluble globular proteins). Difficulties arise in cases where the spectral features of the protein under study do not reflect the characteristics of the spectra within the calibration set. In such situations, an incorrect estimation of the secondary structure is very likely, even though the mathematical treatment of the spectral data is formally correct.

### 3.3.6. Spectral Interference by Amino Acid Side-Chain Absorptions

A common problem for both the curve-fitting method and the pattern-recognition approach is that some amino acid side chains display absorption bands in the amide I/I' spectral region, which, in some proteins, may account for as

much as 15–20% of the total integrated intensity in this region. Sometimes the infrared spectra of amino acids or simple peptides ***(10,16)*** are used for subtraction of the side-chain contributions from the experimental protein spectrum. The spectral parameters of the side-chain absorption bands in the model compounds, however, provide only an approximation because the spectral features of side-chain groups in a protein are influenced by the specific microenvironment of the corresponding group.

### 3.4. Isotope-Edited FTIR Spectroscopy

The assignment of infrared bands to specific groups of a protein can be accomplished by site-specific mutation or by isotopic labeling, and then by comparing the spectra of the unmodified and the modified protein. Site-directed mutagenesis can disturb the structure and function of the protein, thereby complicating the assignments. Isotope substitution has the advantage of being noninvasive, and facilitates band assignment by shifting bands that arise from vibrational modes involving chemical groups, which contain the isotope. Moreover, isotope labeling, such as site-specific $^{13}C$ labeling of the polypeptide backbone, allows FTIR spectroscopy to locate a particular secondary structure within the polypeptide chain and helps analyzing conformational changes that exclusively originate from the labeled site. The most valuable way is site-directed isotope labeling, which is feasible without extra efforts by chemical synthesis of peptides. The biosynthetically incorporation of a site-specifically labeled amino acid in a protein is much more difficult to achieve, and have been reported for only very few proteins yet ***(17)***. What is easier to achieve is uniform labeling of a specific type of amino acid residue in a protein. As an example, we have incorporated $^{13}C$ in the carbonyl position in the polypeptide backbone of all methionine residues in calmodulin. As the replacement of a $^{12}C = O$ group with a $^{13}C = O$ group decreases the amide I' vibration by 35–45 $cm^{-1}$, a comparison of the spectra of the unlabeled and the labeled protein allows the identification of the amide I' bands that originate from the labeled site. In the case of calmodulin, the amide I' band at 1643 $cm^{-1}$ undergoes a small drop in intensity, whereas a weak band assignable to the $^{13}C$-labeled carbonyls appears near 1600 $cm^{-1}$ (compare the solid and dashed line in **Fig. 5A**). Because the nine methionine residues in calmodulin are exclusively located in helices, the observed isotopic shift provides direct evidence for the assignment of the amide I' band at 1643 $cm^{-1}$ to $\alpha$-helical structures present in calmodulin.

Uniformly, $^{13}C$-labeling of a protein completely shifts the amide I' band toward lower wavenumbers (*see* **Fig. 5B**). By mixing a completely $^{13}C = O$-labeled protein with an unlabeled peptide or protein, it then becomes feasible to observe their respective amide I' bands separately. This approach has been used to moni-

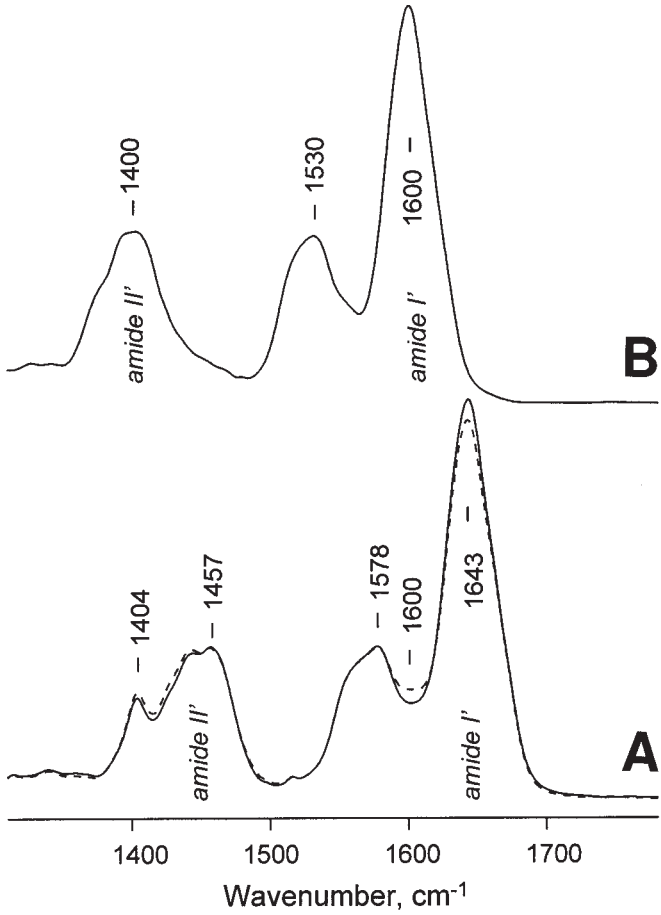


Fig. 5. Amide I' and amide II' regions of the IR spectra of **(A)** unlabeled calmodulin (*solid line*) and calmodulin containing a $^{13}C$ isotope at the carbonyl carbon of all methionine residues (*dashed line*). **(B)** $^{13}C/^{15}N$-uniformly labeled calmodulin. $^{15}N$ labeling produces almost no changes in the amide I' band contour but causes a large shift of the amide II' band of the protein.

tor conformational changes induced by binding of different target peptides to calmodulin ***(18,19)***. It is important to note that, in general, the larger protein (in this case calmodulin, with 148 amino acids) is labeled for over 99% to ensure the lack of any residual intensity from the unlabeled protein. This allows one to detect the much-weaker amide I' bands of the smaller target peptides. The carry-over counterion trifluoroacetic acid used in peptide synthesis, which has a strong infrared band at 1674 $cm^{-1}$, can be very useful as an internal reference to normalize the spectra (*see* **Note 6**).

### 3.5. Time-Resolved FTIR Spectroscopy

Although FTIR spectroscopy has certain limitations in the quantification of protein secondary structure, it represents an excellent tool for monitoring, in relative terms, changes in the conformation of proteins under equilibrium ***(20)*** and nonequilibrium conditions ***(21)***. The highest sensitivity is provided by techniques that can induce the conformational change with minimal sample manipulation and without removing it from the instrument; the most powerful approach involves the use of light to trigger the event. Originally, these techniques were developed for the investigation of light-induced reactions in photoreactive proteins. They can be very specific down to the level of single functional groups in a large molecule with nanosecond time resolution applying sophisticated experimental setups ***(22,23)***. The use of caged compounds creates unique possibilities to initiate reactions by light ***(24)***. For studies of calcium-binding proteins, photolabile derivatives of cation chelating reagents such as DM-nitrophen are of particular interest. DM-nitrophen consist of a photolabile group linked to an EDTA molecule. The EDTA moiety of the molecule is split into two parts upon photolysis by a UV flash, which triggers the release of calcium in the sample under study. The time resolution is only limited by the intrinsic reaction rates of the cage and diffusion to the protein. The interpretation of the spectra is complicated by the fact that the spectral features of the caged compound itself change after photolysis and overlap with some protein bands of interest. Using caged $Ca^{2+}$, ATP, or ADP, function-related events on a molecular level in the sarcoplasmic reticulum $Ca^{2+}$-ATPase were analyzed by time-resolved Fourier transform infrared difference spectroscopy ***(25–27)***.

## 4. Notes

1. It should be noted that commercial suppliers of FTIR spectrometers provide instruments that plot spectra either with increasing or decreasing wavenumber.
2. The potential problems that can arise when using $CaF_2$ cells in infrared measurements of calcium-binding proteins have been descrived in detail elsewhere ***(28)***.
3. Altering the solvent from $H_2O$ to $D_2O$ needs to be considered carefully as a potential experimental variable. Before an IR experiment is done, it is often advisable to completely exchange all protein amide groups, so that no further changes occur in the amide II band in the course of the measurement. Some internal backbone NH protons in certain proteins can exchange rather slowly and may require extended incubation in $D_2O$ at room temperature, or treatment at higher temperature, or other extreme conditions (e.g., higher pH). Nuclear magnetic resonance spectroscopy measurements indicate that at 10 C below the proteins denaturation temperature hydrogen exchange is usually very rapid, while minimizing protein denaturation (*see* also **Subheading 3.2.1.**). Note that amide hydrogen exchange is markedly dependent on pH, with low pH slowing down the reaction.

4. ATR-FTIR spectroscopy is, of course, an important technique for studying the association of membrane proteins or peptides with membranous or membrane minetic surfaces. The technique allows detection of the changes in secondary structure that accompany their binding; moreover the orientation (e.g., parallel or perpendicular to the membrane surface) can be determined. (e.g., **ref. *29***).
5. Protein concentrations are most-accurately determined spectroscopically using published extinction coefficients or by quantitative amino acid analysis of stock solutions.
6. The trifluoro acetate has a strong infrared absorption band at 1673 $cm^{-1}$ that can overlap with the amide I band of the peptide. In some synthetic peptide preparations, it is best to remove the TFA completely. This can be accomplished by repeated lyophilization from 10 m*M* hydrochloric acid ***(30)***.

## References

1. Goormaghtigh, E., Cabiaux, V., and Ruyschaert, J.-M. (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy, in *Subcellular Biochemistry,* vol. 23, *Physicochemical Methods in the Study of Biomembranes* (Hilderson, H. J. and Ralston, B. G., eds.), Plenum, New York, pp. 329–450.
2. Fabian, H. and Mantsch, H. H. (1995) Ribonuclease A revisited: Infrared spectroscopic evidence for the lack of native-like structures in the thermally denatured state. *Biochemistry* **33,** 10,725–10,730.
3. Alben, J. O. and Fiamingo, F. G. (1984) Fourier transform infrared spectroscopy, in *Optical Techniques in Biological Research* (Rousseau, D. L., ed.), Academic, New York, pp. 133–179.
4. Venyaminov, S. Y. and Prendergast, F. G. (1997) Water ($H_2O$ and $D_2O$) molar absorptivity in the 1000–4000 $cm^{-1}$ range and quantitative infrared spectroscopy of aqueous solutions. *Anal. Biochem.* **248,** 234–245.
5. Jackson, M. and Mantsch, H. H. (1995) The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30,** 95–120.
6. Nara, M., Tasumi, M., Tanokura, M., Hiraoki, T., Yazuwa, M., and Tsutsumi, A. (1994) Infrared studies of interaction between metal ions and $Ca^{2+}$-binding proteins. Marker bands for identifying the types of coordination of the side-chain $COO^-$ groups to metal ions in pike parvalbumin. *FEBS Lett.* **349,** 84–88.
7. Nara, M., Tanokura, M., Yamamoto, T., and Tasumi, M. (1995) A comparative study of the binding effects of $Mg^{2+}$, $Ca^{2+}$, $Sr^{2+}$, and $Cd^{2+}$ on calmodulin by Fourier-transform infrared spectroscopy. *Biospectroscopy* **1,** 47–54.
8. Moffatt, D. J. and Mantsch, H. H. (1992) Fourier resolution enhancement of infrared spectral data. *Methods Enzymol.* **210,** 192–200.
9. Surewicz, W. K., Mantsch, H. H., and Chapman, D. (1993) Determination of protein secondary structure by Fourier transform infrared spectroscopy. *Biochemistry* **32,** 389–394.

10. Chirgadze, Y. N., Fedorov, O. V., and Trushina, N. P. (1975) Estimation of amino acid residue side-chain absorption in the infrared spectra of protein solutions in heavy water. *Biopolymers* **14,** 679–694.
11. Byler, D. M. and Susi, H. (1986) Examination of the secondary structure of proteins by deconvoluted FT-IR spectra. *Biopolymers* **25,** 469–487.
12. Reisdorf, W. C. and Krimm, S. (1996) Infrared Amide I' band of the coiled coil. *Biochemistry* **35,** 1383–1386.
13. Trewhella, J., Liddle, W. K., Heidorn, D. B., and Strynadka, N. (1989) Calmodulin and troponin C structures studied by Fourier transform infrared spectroscopy: effects of $Ca^{2+}$ and $Mg^{2+}$ binding. *Biochemistry* **28,** 1294–1301.
14. Jackson, M., Haris, P. I., and Chapman, D. (1991) Fourier transform infrared spectroscopic studies of $Ca^{2+}$-binding proteins. *Biochemistry* **30,** 9681–9686.
15. Fabian, H., Yuan, T., Vogel, H. J., and Mantsch, H. H. (1996) Comparative analysis of the amino- and carboxy-terminal domains of calmodulin by Fourier transform infrared spectroscopy. *Eur. Biophys. J.* **24,** 195–201.
16. Venyaminov, S. Y. and Kalnin, N. N. (1990) Quantitative IR spectrometry of peptide compounds in water ($H_2O$) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers* **30,** 1243–1257.
17. Ludlam, C. F. C., Sonar, S., Lee, C.-P., Coleman, M., Herzfeld, J., RajBhandary, U., and Rothschild, K. J. (1995) Site-directed isotope labeling and ATR-FTIR difference spectroscopy of bacteriorhodopsin: the peptide carbonyl group of Tyr 185 is structurally active during the bR N transition. *Biochemistry* **34,** 2–6.
18. Zhang, M., Fabian, H., Mantsch, H. H., and Vogel, H. J. (1994) Isotope-edited FTIR spectroscopy studies of calmodulin's interaction with its target peptides. *Biochemistry* **33,** 10,883–10,888.
19. Yuan, T., Walsh, M. P., Sutherland, C., Fabian, H., and Vogel, H. H. (1999) Calcium-dependent and -independent interactions of the calmodulin-binding domain of cyclic nucleotide phopshodiesterase with calmodulin. *Biochemistry* **38,** 1446–1455.
20. Fabian, H., Schultz, C., Backmann, J., Saenger, W., Mantsch, H. H., and Naumann, D. (1994) Impact of point mutations on the structure and thermal stability of ribonuclease T1 in aqueous solution probed by Fourier transform infrared spectroscopy. *Biochemistry* **33,** 10,725–10,730.
21. Reinstädler, D., Fabian, H., Backmann, J., and Naumann, D. (1996) Refolding of thermally and urea denatured ribonuclease A monitored by time-resolved FTIR spectroscopy. *Biochemistry* **35,** 15,822–15,830.
22. Mäntele, W. (1993) Reaction-induced infrared difference spectroscopy for the study of protein function and reaction mechanisms. *Trends Biochem. Sci.* **18,** 197–202.
23. Siebert, F. (1996) Equipment: slow and fast infrared kinetic studies, in *Infrared Spectroscopy of Biomolecules* (Mantsch, H. H. and Chapman, D., eds.), Wiley, New York, pp. 83–106.
24. Cepus, V., Ulbrich, C., Allin, A. T., and Gerwert, K. (1998) Fourier transform infrared photolysis studies of caged compounds. *Methods Enzymol.* **291,** 223–245.

25. Georg, H., Barth, A., Kreutz, W., Siebert, F., and Mäntele, W. (1994) Structural studies of sarcoplasmic reticulum $Ca^{2+}$-ATPase upon $Ca^{2+}$ binding studied by simulteneous measurement of infrared absorbance changes and changes of intrinsic protein fluorescence. *Biochim. Biophys. Acta* **1188,** 139–150.
26. Troullier, A., Gerwert, K., and Dupont, Y. (1996) A time-resolved Fourier transform infrared difference spectroscopy study of the sarcoplasmic reticulum $Ca^{2+}$-ATPase: kinetics of the high-affinity calcium binding at low temperature. *Biophys. J.* **71,** 2970–2983.
27. Barth, A., Kreutz, W., and Mäntele, W. (1997) $Ca^{2+}$ release from the phosphorylated and the unphosphorylated sarcoplasmic reticulum $Ca^{2+}$-ATPase results in parallel structural changes. *J. Biol. Chem.* **272,** 25,507–25,510.
28. Moncrieffe, M. C., Venyaminov, S. Y., and Prendergast, F. G. (1999) A pitfall in the use of calcium fluoride cells for infrared spectroscopic measurements of calcium-binding proteins. *Anal. Biochem.* **268,** 163–164.
29. Skaron, M., Oren, Z., Shai, Y., and Anglister J. (1999) 2D-NMR and ATR-FTIR study of the structure of a cell-selective diastereomers of melittin and its orientation in phospholipids. *Biochemistry* **38,** 15,305–15,316.
30. Lewis, R. N. A. H., Prenner, E. J., Kondejewski, L. H., Flach, R., Mendelsohn, R., Hodges, R. S., and McElhaney, R. N. (1999) FTIR spectroscopic studies of the interaction of the antimicrobial peptide gramicidin S with lipid micelles and with lipid monolayer and bilayer membranes. *Biochemistry* **38,** 15,193–15,203.

# 6

# Steady-State Fluorescence Spectroscopy

**Aalim M. Weljie and Hans J. Vogel**

## 1. Introduction

Fluorescence spectroscopy has long been a popular method for protein studies from which researchers have garnered a wealth of biophysical information ***(1,2)***. Several specific fluorescence methods have been recently well reviewed ***(3–5)*** and readers are encouraged to seek out these references for theory and methods complimentary to those presented in this chapter. The basic selling features for the general use of this tool in biological systems include the relatively low concentrations of sample material required, the occurrence of natural fluorophores in proteins such as tryptophan and tyrosine, the breadth of fluorescence experiments available, and the comparatively simple (and inexpensive) equipment required for most experiments. It is no surprise then that the literature is replete with examples of calcium-binding proteins which have in one way or another been characterized by some fluorescence method. Information available to the researcher includes, but is not limited to, biochemical characteristics such as conformational changes, protein–protein interactions, metal-binding information, membrane localization, long-range distance measurements, and kinetic/dynamic parameters. The majority of this chapter will concentrate on protocols for simple steady-state single-tryptophan fluorescence measurements to probe protein–peptide interactions. References to other fluorescence methods and applications will also be provided.

Tryptophan is a popular intrinsic protein fluorescence probe (fluorophore) because of the high sensitivity of the indole moiety to electronic excitation. Generally, fluorophores have extended π systems or significant electron density (such as that in lanthanide metal ions) that allow for electronic transitions to high-energy excited states upon absorption of photon radiation. The fluorophore is excited by a particular wavelength of light, or more correctly

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

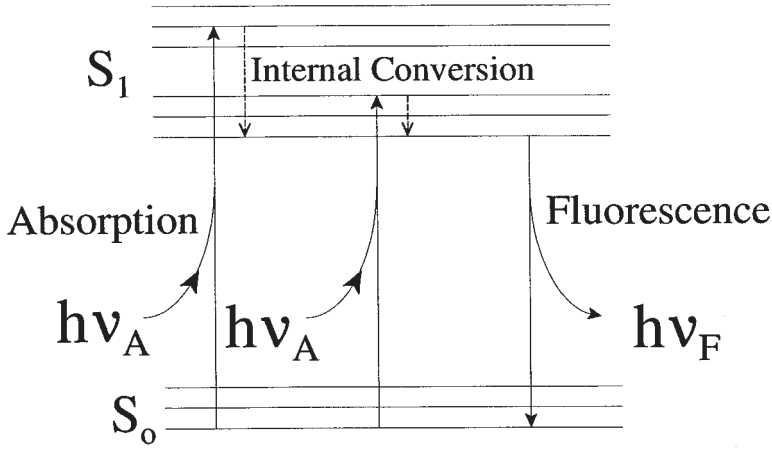


Fig. 1. Simplified Jablonski diagram depicting absorption of photon energy ($h\nu_A$) from the ground state singlet ($S_o$). The result is a series of excited-state singlet states ($S_1$), which undergo internal conversions to the lowest $S_1$ state. Of the possible relaxation pathways available back to $S_o$, one possibility is the emission of photon radiation ($h\nu_F$), which gives rise to the phenomenon of fluorescence. Adapted from **ref. *1***.

over a range of wavelengths, as can be demonstrated by an excitation spectrum. Fluorescence emission occurs when the electronic excited state of the fluorophore returns to the ground state by emitting photons, forgoing other relaxation pathways that are nonradiative, such as conversion to heat. The resulting fluorescence emission occurs at longer wavelengths than that of excitation because of energy loss as a result of microrelaxation in the excited state (*see* **Fig. 1**).

The power of fluorescence lies in the ability to probe the environment immediately surrounding the fluorophore of interest. There are numerous factors that affect the electronic excited state and the subsequent fluorescence of a particular fluorophore, such as hydrophobicity, viscosity, and mobility. Readers are referred to other excellent discussions of fluorescence theory ***(1,2,6,7)*** for in-depth treatment of these physical phenomenon. In many cases, these effects can be observed with basic fluorescence instrumentation using the methods described below (*see* **Subheading 3.**).

An example of a major application of fluorescence to calcium-binding proteins is demonstrated in probing the interaction of calcium-sensitive signaling proteins, such as calmodulin (CaM) or troponin C, with binding targets in the presence and absence of calcium ***(8–12)***. The intrinsic tyrosine residues of these proteins exhibit marked fluorescence intensity changes when calcium binds, as shown in **Fig. 2** with CaM. This phenomenon is extremely sensitive to calcium

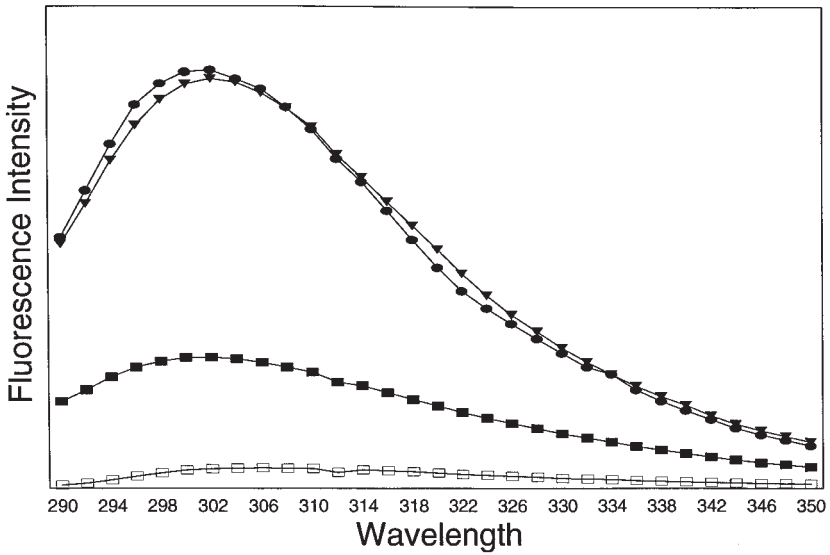


Fig. 2. Steady-state tyrosine fluorescence spectra of CaM and a synthetic peptide encompassing the CaM-binding domain of constitutive nitric oxide synthase (cNOS). The $Ca^{2+}$–CaM tyrosine fluorescence (*filled ovals*) is markedly intense compared with the apo- form of the protein (*filled squares*). The cNOS peptide alone (*open squares*) shows no tyrosine fluorescence, and the CaM–cNOS complex (*filled triangles*) fluorescence intensity is very similar to that of $Ca^{2+}$–CaM alone (Weljie and Vogel, unpublished observations).

concentrations. The tyrosine fluorescence is unchanged again when a synthetic peptide encompassing a CaM-binding domain is added to $Ca^{2+}$–CaM. In spite of these transitions, applications for tyrosine fluorescence are somewhat limited, as the presence of tryptophan precludes exclusive tyrosine excitation. However, in spite of the fact that CaM and troponin C do not contain tryptophan residues, tryptophan fluorescence has been shown to be an important tool for probing intermolecular interactions. For example, our lab has studied the interaction of CaM with synthetic peptides encompassing the CaM-binding domains from various CaM target proteins. The tryptophan in many of these synthetic target peptides acts as excellent intrinsic fluorescent probes. Binding of a target peptide to wild-type CaM is accompanied by a significant increase in tryptophan fluorescence intensity and a marked blueshift in the maximum-fluorescence emission peak (*see* **Fig. 3**). Altering the chemical nature of the CaM side chains produces markedly different spectra, such as decreased fluoresence intensity when selenomethionine residues replace methionine, demonstrating the rapid use of fluorescence to provide useful information.

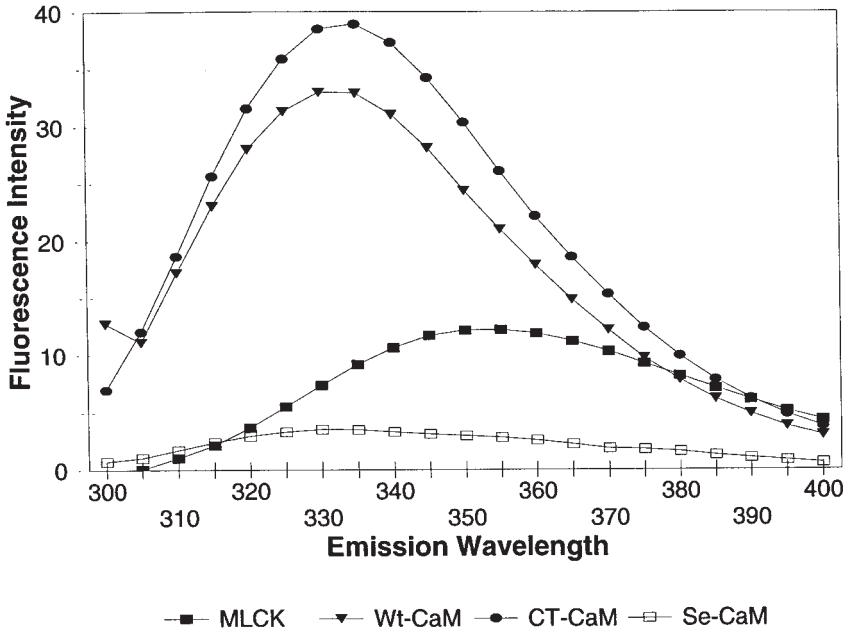


Fig. 3. Steady-state tryptophan fluorescence spectra of a synthetic peptide encompassing the CaM-binding domain of skeletal myosin light chain kinase alone (*filled rectangles*), and in complexes with wild-type CaM (*filled triangles*), a mutant calmodulin where the *C*-terminal methionine residues were replaced with leucine (*filled ovals*), and a CaM where the methionine residues were replaced by selenomethionine. Note that there is a blue-shift in the observed peak intensity of tryptophan fluorescence once the fluorophore is sequestered into the hydrophobic binding pocket. Also, changing the envrionment of the tryptophan by altering the chemical nature of the CaM side chains produces remarkable intensity variations ***(10,12)***.

Once such a complex is formed and the peak fluorescence known, one can establish the degree of solvent exposure via quenching experiments. Fluorescence is said to be quenched when a species proximal to the fluorophore provides an alternate relaxation pathway for the excited electronic state to return to the ground state. Common quenching agents include potassium iodide, cesium chloride, and acrylamide. The degree of quenching is generally given using a Stern-Volmer plot, from which one obtains Stern-Volmer quenching constants ($K_{SV}$) (*see* **Fig. 4**).

Detailed below is a method for acquiring simple tryptophan emission spectra, and subsequently obtaining Stern-Volmer quenching constants. Selected examples of other applications of fluorescence to calcium-binding proteins examples are provided in **Subheading 4.2.**

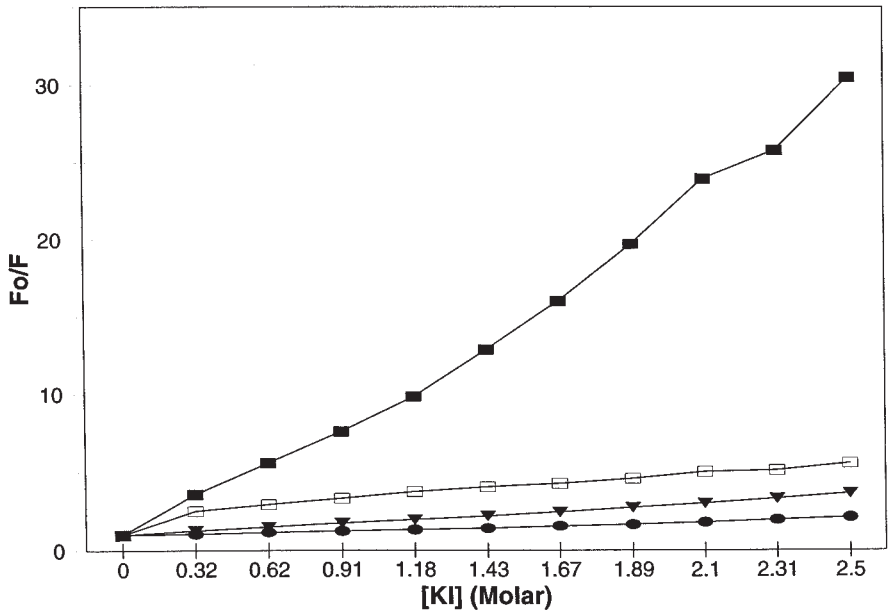


Fig. 4. Potassium iodide Stern-Volmer quenching plots of a synthetic peptide encompassing the CaM-binding domain of skeletal myosin light-chain kinase alone (*filled rectangles*), and in complexes with wild-type calmodulin (*filled triangles*), a mutant calmodulin where the *C*-terminal methionine residues were replaced with leucine (*filled ovals*), and a CaM where the methionine residues were replaced by selenomethionine. Notice that the free peptide shows a much greater degree of solvent exposure than the CaM-target complexes as evidenced by the increased slope ***(10,12)***. The slope in a Stern-Volmer quenching plots provides the Stern-Volmer quenching constant $K_{SV}$.

## 2. Materials

1. Wavelength-scanning spectrofluorimeter capable of excitation at 295 nm (*see* **Note 1**) and detection of fluorescence emission between 300 and 450 nm, such as the Hitachi F-2000. Ideally the instrument should either be connected to a microcomputer or have data transfer capability to facilitate subsequent analysis of the data.
2. Quartz cells that have two transparent faces; one for excitation and one for emission. Generally these faces will be at 90° to each other and cells are available as rectangular cuvets with four clear sides. For cells with a path length of 1.0 cm, the sample volume needed is approx 3.0 mL.
3. Standard buffers chosen to meet pH and ionic requirements of the protein(s) under study, bearing in mind that absorption of buffer components in the UV range is undesirable. The absorption of the buffer can be measured using a standard UV/

VIS spectrophotometer and acquiring an absorption spectrum between 250 and 320 nm. It may also be worthwhile to acquire a UV/VIS spectrum of the buffer between 300 and 450 nm to ensure that there is no transfer of tryptophan-fluorescence emission to the buffer. If one is examining the interaction of charged species, the addition of 50–100 m*M* KCl or NaCl may be beneficial to prevent nonspecific interactions. Finally, phosphate buffers are to be avoided in studies involving $Ca^{2+}$ because of the precipitation of calcium phosphate solid.

4. For quenching studies, a saturated stock solution of the chosen quenching agent is required. For example, a stock solution of potassium iodide can be made by adding 30.0 g of KI to 21.0 mL of water, to form 30.0 mL of a 6.0 *M* solution (*see* **Note 2**).

## 3. Method

### 3.1. Sample Preparation

1. Purified, dry protein is weighed and then dissolved in the chosen buffer. Care must be taken to ensure that the pH does not vary significantly. This is especially necessary for calcium-binding proteins which often have significant numbers of acidic residues in the calcium binding regions (*see* **Note 3**).
2. The concentration of the sample must be quantitatively determined, as the dry mass is not a reliable indicator of protein mass. The method of choice for such concentration determinations of proteins (or peptides) without tryptophan and tyrosine is quantitative amino acid analysis. If the protein sequence is known, the absorbance at 280 nm can be used if the protein contains tryptophan or tyrosine residues. The extinction coefficient can be calculated as:

$$\varepsilon_{280} = x\,\varepsilon_{tyr} + y\,\varepsilon_{trp} + z\,\varepsilon_{cys}$$

where $x$, $y$, and $z$ are the numbers of each residue in the sequence, and the value of $\varepsilon_{tyr}$ is 1280 $M^{-1}cm^{-1}$, $\varepsilon_{trp}$ is 5690 $M^{-1}cm^{-1}$, and $\varepsilon_{cys}$ (cystine) is 120 $M^{-1}cm^{-1}$. ***(13)*** (*see* **Note 4**).

3. If a number of samples are to be made of the same protein or peptide, a 100X stock solution can be made, and then a sample pool made with a volume slightly greater than the total volume of all samples. For example, if 10 3.0 mL samples of protein are needed, then a 30.5-mL pool might be made, and each sample made from this pool. This will help ensure that the concentration of each sample remains consistent, which is especially important when comparing spectra of the same species under different conditions. Ensure that the volume used also takes into account other reagents that might be added, such as calcium stock solutions (*see* **Note 5** for suggested stock concentrations).
4. Samples involving binding (calcium and or protein/peptide ligands) should be allowed to come to a complete equilibrium prior to spectrum acquisition. For stable species, samples can be prepared the day prior to use, and then stored at 4° overnight. Ensure that the samples are allowed to equilibrate to room tempera-

ture for 15–30 min prior to fluorescence analysis. For samples that are temperature sensitive, many spectrometers can be outfitted with a jacketed cuvet holder to control sample temperature during spectral acquisition.

### 3.2. Instrument Setup and Calibration

1. Turn instrument on several minutes to half an hour prior to experimentation to allow the light source and electronics to stabilize (check specifications for individual fluorimeters).
2. For tryptophan fluorescence, set the excitation wavelength to 295 nm (*see* **Note 1**), and set values for the excitation and emission bandpass shutters. These values should be around 1.5–5 nm for excitation, although certain instruments do not allow for bandpass sizes less than 10 nm, and 4–10 nm for emission intensities (*see* **Note 6**).
3. Depending on the lamp being used in the fluorimeter, the fluorescence intensity may decrease with the age of the lamp. If this is suspected to be a factor, the fluorescence intensity can be normalized to a reference sample (*see* **Note 7**).

### 3.3. Obtaining Fluorescence Emission Spectrum

1. Ensure that the sample cell is clean and dry prior to use to prevent extraneous water from changing the sample concentration.
2. Obtain a baseline fluorescence-emission scan of buffer alone from 300 to 450 nm keeping the excitation constant at 295 nm. A typical scan rate would be between 25–100 nm/min. A buffer baseline correction should also be performed with other species present in the sample, but which do not contain tryptophan, such as other proteins/ligands.
3. Use the same parameters to obtain spectra of the fluorescence sample(s). Ensure that the sample cell is washed and dried thoroughly between acquisitions (*see* **Note 8**).
4. Subtract the appropriate baseline spectrum from the spectra of interest to obtain the final emission spectrum.

### 3.4. Quenching Studies

1. Use the method outlined in **Subheading 3.3.** to obtain an emission spectrum of the sample. Note the wavelength of peak-fluorescence emission, and the fluorescence intensity at this point.
2. Set the fluorimeter-emission wavelength detection to the aforementioned emission maximum wavelength.
3. Add an aliquot of quenching agent directly to the sample in the cuvet, and mix thoroughly. As a simple example, with a 3.0-mL initial sample volume, 50-μL additions of 6.0 *M* KI will provide reasonable results for a species in which the fluorophore is well protected from the solvent (*see* **Note 2** for more information on quenching agents).

4. A series of emission intensities is obtained at various quenching agent concentrations by incrementally adding the same amount of quenching agent (e.g., 50 μL in the above example to a final quenching concentration of 2.5 *M*).
5. The observed intensities must be corrected for dilution effects by multiplying the observed intensity at each point by the dilution factor ($V/V_o$ where $V$ is the volume at a given point, and $V_o$ is the initial volume without any quenching agent present). For example, if 50 μL quenching agent is added to a 3.0-mL sample, the resultant fluorescence intensity must be multiplied by (3.05/3.0).
6. The Stern-Volmer quenching constants can be derived from a plot that follows the following equation:

$$F_o/F = 1 + K_{SV}[Q]$$

where $F_o$ is the fluorescence intensity without any quenching agent present, and the values of $F$ are the fluorescence intensities at given concentrations of the quenching agent Q. The slope of the plot of $F_o/F$ vs $[Q]$ will be the Stern-Volmer quenching constant, $K_{SV}$ ***(1)***.
7. Ensure that changes in fluorescence intensity upon dilution are not a result of concentration changes by repeating **Subheading 3.4.**, **steps 4** and **5** with water alone as a control.

## 4. Notes

### 4.1. Emission Spectra and Quenching of Tryptophan Fluorescence

1. The optimal excitation wavelength is dependent on the combination of fluorophores present in the sample of interest. Tryptophan is generally excited at 295 nm for proteins in order to minimize tyrosine fluorescence, which is maximally excited at 278 nm. This wavelength provides sufficient excitation to observe significant signal at low micromolar protein concentrations on an inexpensive spectrofluorimeter. Lower protein concentrations (up to nanomolar) are feasible as the detection systems become more complex (and expensive). An appropriate excitation wavelength can be determined by running a series of excitation spectra on isolated fluorophores. These profiles will provide an indication as to which wavelengths provide mutual excitation, and more importantly, where individual fluorophores can be selectively excited. Often, this wavelength will not be where a fluorophore exhibits it's peak extinction coefficient, hence, there may be a trade-off between selectivity and sensitivity.
2. The choice of quenching agent is dependent on the system under study and tolerable dilution effects. We have used KI, CsCl, and acrylamide successfully for quenching of calmodulin-target peptide complexes with equivalent results with 100 m*M* salt in the buffer. In each of these cases, the quenching agent has a different charge ($I^-$, $Cs^+$, acrylamide is polar, but neutral), and different quenching efficiencies. Acrylamide is a very efficient quenching agent, hence smaller dilution effects will be observed if this is a concern, however caution must be used as it is a neurotoxin. Also, corrections must be made for the absorption of

acrylamide. KI is sensitive to light degradation; therefore, solutions should be made fresh and care taken with handling.

3. Alternatively, the sample can be exchanged from a different buffer into the fluorescence buffer by passing the sample through a size exclusion column containing G-25 resin (Pharmacia).
4. There are also several sites on the World Wide Web where one can obtain estimates of protein extinction coefficients using more complex factors. For example, the ProtParam tool estimates $\varepsilon_{280}$ based on both fully reduced *cys* residues, or on complete disulphide bond formation, in addition to tryptophan and tyrosine contributions.
5. In preparation of samples, the following concentration guidelines from our lab might be useful. Protein samples containing the fluorophore of interest are used at a concentration of 10–12 μ*M*. Any other protein species binding to this one is added to a 10–20% molar excess. For addition of $Ca^{2+}$ ions, a 50-m*M* stock of $CaCl_2$ is added to a final concentration of 1 m*M*. Calcium-free studies are ensured by the addition of a 500 m*M* stock of EDTA to a final concentration of 5 m*M*.
6. The optimal bandpass settings are dependent on numerous technical factors including the optics of the instrument, the signal to noise ratio of the photomultiplier tube, and the signal from the sample, all of which is unique to a given setup. Generally speaking, as one increases the excitation/emission bandpass sizes, more electronic transitions are excited/detected simultaneously; hence, fine structure is lost at larger values. It is important to note that the bandpass value is taken as the full slitwidth, with the chosen wavelength at the center. For example, a bandpass of 5 nm for excitation at 295 implies light from 292.5 to 297.5 nm would pass through, unlike the definition used in other spectroscopic methods (i.e., 295 +/– 5 nm). If the intensity of the fluorescence signal being monitored shows small changes with mediocre signal to noise, the emission bandpass can be increased (to 50 nm for example) in order to provide the equivalent of an integral of the emission peak.
7. Several methods of referencing can be used for normalization and calibration of the fluorimeter. The best trade-off between simplicity and accuracy is to use a chemical sample of known composition and concentration (such as 5 μ*M* tryptophan at pH 7.0) before each fluorescence session and normalize the acquired data to this reference.
8. An ethanol or detergent wash followed by a minimum of 10 rinsings with distilled water should be performed between samples. For smaller cells, an air aspirator will facilitate the cleaning and drying process.

### 4.2. Other Fluorescence Methods and Applications

9. After having reviewed the literature and performed initial fluorescence experiments on a (suspected) calcium-binding protein, here are several simple experiments suggested for initial analysis:
   a. Excitation and emission spectra of protein alone, and then +/– $Ca^{2+}$ titration.

**Table 1**
**Advanced Fluorescence Methods and Selected Applications**

| Protein/ Fluorophore | Brief method details | Application | Ref. |
|---|---|---|---|
| Parvalbumin, oncomodulin, calmodulin/$Tb^{3+}$ | FRET from aromatic amino acids in the $Ca^{2+}$- binding loops of the proteins to bound $Tb^{3+}$. | Use of $Tb^{3+}$ as a sensitive luminescent probe of the structure and function of EF-hand $Ca^{2+}$-binding loops | *(15)* |
| Calmodulin, troponin C/tryptophan, DANSYL | Frequency domain FRET used to probe the distance between the DANSYLated *N*-terminus and Trp of melittin. These measurements were performed for the free peptide as well as in complexes with CaM, TnC, and in vesicles. | Determine the distribution of distances present in solution between the two fluorophores. The result can be interpreted as the degree of conformational freedom. | *(16)* |
| Calmodulin/tryptophan | Steady state fluorescence of CaM complexes with CaM-target peptides. CaM proteins were engineered such that methionine residues were replaced with unnatural amino acid analogs. | Probing the role of methionine side-chains in the sequestering of CaM-target peptides. | *(10,12)* |
| Calmodulin/tyrosine | FRET between two tyrosines by using a nitro-tyrosine derivative as the acceptor, and a normal tyrosine as the donor. | Determination of distance separation in both the apo and $Ca^{2+}$-saturated states. | *(17)* |
| 68Calmodulin/GFP | Distinctly colored mutant Green Fluorescent proteins are genetically appended to the termini of a CaM/CaM-target chimera and monitored for fluorescence resonance energy transfer. | Detection of localized concentrations of $Ca^{2+}$. Can be targeted to specific organelles within living cells. | *(18)* |
| Phospholipid scramblase/tryptophan, $Tb^{3+}$ | FRET evidence that there is coordinate metal ion binding from protein trp to $Tb^{3+}$ in an EF- hand-like domain. The $Tb^{3+}$ is competed out by $Ca^{2+}$. | Confirmation that the EF-hand looplike segement contributes directly to $Ca^{2+}$-binding. | *(19)* |
| C2 domain of Cytosolic Phospholipase $A_2$/fluorescin | Cysteine scanning mutagenesis used to attach a fluorescin probe in 16 locations. $Ca^{2+}$ triggered environmental changes were probed based on intensity changes and Stern-Volmer quenching constants. | Indication that the membrane docking surface of the C2 domain is localized to the same surface that binds a pair of $Ca^{2+}$ ions. | *(20)* |
| Annexin V/fluorescin, NBD | Fluorescence recovery after photobleaching experiments employed to probe the concentration effects of annexin V on its lateral mobility in a mixed lipid system. | Demonstration that the annexin is hindered by its specific interaction with one type of lipid; localization of this interaction to the headgroup. | *(21)* |

b. Quenching studies with several quenching agents to obtain $K_{SV}$, and then repeat the quenching experiments at several $Ca^{2+}$ ion concentrations to obtain a plot of $K_{SV}$ vs $[Ca^{2+}]$. If the calcium-binding protein interacts with another species containing a fluorophore, such as a tryptophan containing protein/peptide, or lanthanide ion:
c. Repeat the excitation and emission spectra and quenching experiments with all species varied (e.g., protein alone, peptide alone, protein + $Ca^{2+}$, peptide + $Ca^{2+}$, protein + peptide, protein + peptide + $Ca^{2+}$). If the secondary binding species is a lanthanide ion that binds in the $Ca^{2+}$-binding region of the protein, then fluorescence of both the protein and the ion can be monitored in a competitive titration experiment.
d. Red-edge emission spectral effect (REES) spectroscopy can be used to monitor different populations of Trp ***(14)***, and determining if this is a sensitive tool for following $Ca^{2+}$-binding in titrations.
e. Fluorescence energy transfer (or Förster's resonance effect spectroscopy, FRET) can be used to monitor the distance between two specially chosen fluorophores between 10–75 Å apart ***(3)***. There is a high possibility that labeling with an extrinsic fluorophore will be necessary ***(5)***. Once a suitable system is established, quenching experiments can again be used to determine the relative solvent exposures of the two fluorophores, with and without $Ca^{2+}$.
f. Fluorescence anisotropy is another tool used to study the interaction between two biomolecules, and has successfully been applied to calcium binding systems ***(4)***.

10. A selected set of examples that encompass major fluorescence applications to calcium-binding proteins is given in **Table 1**. Many of these applications are dependent on specific fluorophores (e.g., unique Trp residues, or attachment sites for extrinsic fluorophores). The methods described in **Table 1** also often require instrumentation and expertise beyond the scope of this chapter, but are provided for a demonstration of the usefulness of this technique.
11. $Tb^{3+}$ is a popular fluorescent $Ca^{2+}$ analogue, which can be used both intrinsically as an indicator for metal ion-binding, $Ca^{2+}$-binding competition experiments, and also as part of a donor/acceptor pair in resonance transfer analysis. Caution should be used in assessing structure–function relationships of proteins with bound $Tb^{3+}$ however, as protein activity can be effected by this substitution.
12. Mutagenesis studies in which Tyr/Phe residues are replaced by Trp for fluorescence have been reported. It has been shown however that in the case of troponin C, such mutagenesis alters the $Ca^{2+}$-binding properties of the protein, and results using such mutants must therefore be carefully evaluated ***(22,23)***.

## Acknowledgments

We would like to offer our sincere gratitude to R. J. Turner for insightful discussions and suggestions. A. M. Weljie would like to thank the National Sciences and Engineering Research Council and the Alberta Heritage Fund for Medical Research (AHFMR) for support. Also, H. J. Vogel is an AHFMR Scientist.

## References


1. Lackowicz, J. R. (1983) *Principles of Fluorescence Spectroscopy*. Plenum, New York.
2. Permyakov, E. A. (1993) *Luminescent Spectroscopy of Proteins*. CRC, Boca Raton, Florida.
3. Selvin, P. R. (1995) Fluorescence resonance energy transfer. *Methods Enzymol.* **246,** 300–334.
4. Jameson, D. M. and Sawyer, W. H. (1995) Fluorescence anisotropy applied to biomolecular interactions. *Methods Enzymol.* **246,** 283–300.
5. Waggoner, A. (1995) Covalent labeling of proteins and nucleic acids with fluorophores. *Methods Enzymol.* **246,** 362–373.
6. Chen, Y. and Barkley, M. D. (1998) Toward understanding tryptophan fluorescence in proteins. *Biochemistry* **37,** 9976–9982.
7. Callis, P. R. (1997) 1La and 1Lb transitions of tryptophan: applications of theory and experimental observations to fluorescence of proteins. *Methods Enzymol.* **278,** 113–150.
8. O'Neil, K. T., Wolfe, H. R., Jr., Erickson-Viitanen, S., and DeGrado, W. F. (1987) Fluorescence properties of calmodulin-binding peptides reflects alpha-helical periodicities. *Science* **236,** 1454–1456.
9. Chabbert, M., Piemont, E., Prendergast, F. G., and Lami, H. (1995) Fluorescence of a tryptophan bearing peptide from smooth muscle myosin light chain kinase upon binding to two closely related calmodulins. *Arch. Biochem. Biophys.* **322,** 429–436.
10. Yuan, T., Weljie, A. M., and Vogel, H. J. (1998) Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: orientation of peptide and protein binding. *Biochemistry* **37,** 3187–3195.
11. Yuan, T. and Vogel, H. J. (1998) Calcium-calmodulin-induced dimerization of the carboxyl-terminal domain from petunia glutamate decarboxylase. A novel calmodulin-peptide interaction motif. *J. Biol. Chem.* **273,** 0328–0335.
12. Weljie, A. M. and Vogel, H. J. (1999) Tryptophan fluorescence of calmodulin binding domain peptides interacting with calmodulin containing unnatural methionine analogues. *Protein Eng.*, in press.
13. Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319.
14. Demchenko, A. P. and Ladokhin, A. S. (1988) Red-edge-excitation fluorescence spectroscopy of indole and tryptophan. *Eur. Biophys. J.* **15,** 369–379.
15. Hogue, C., MacManus, J. P., Banville, D., and Szabo, A. G. (1992) Comparison of Terbium(III) luminescence enhancement in mutants of EF hand calcium binding proteins. *J. Biol. Chem.* **267,** 13,340–13,347.
16. Lakowicz, J. R., Gryczynksi, I., Laczko, G., Wiczk, W., and Johnson, M. L. (1994) Distribution of distances between the tryptophan and the N-terminal residue of melittin in its complex with calmodulin, Troponin C, and phospholipids. *Protein Sci.* **3,** 628–637.

17. Steiner, R. F. and Motevalli-Alibadi, M. (1984) The determination of the separation of Tyrosine-99 and Tyrosine-138 in calmodulin: radiationless energy transfer. *Arch. Biochem. Biophys.* **234,** 522–530.
18. Miyawaki A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y. (1997) Fluorescent Indicators for $Ca^{2+}$ based on Green Fluorescent *Proteins* and Calmodulin. *Nature* **388,** 882–887.
19. Stout, J. G., Zhou, Q., Wiedmer, T., and Sims, P. J. (1998) Change in conformation of plasma membrane phospholipid scramblase induced by occupancy of its $Ca^{2+}$ binding site. *Biochemistry* **37,** 14,860–14,866.
20. Nalefski, E. A. and Falke, J. J. (1998) Location of the membrane-docking face on the $Ca^{2+}$-activated C2 domain of cytosolic phospholipase $A_2$. *Biochemistry* **37,** 17,642–17,650.
21. Cezanne, L., Lopez, A., Loste, F., Parnaud, G., Saurel, O., Demange, P., and Tocanne, J.-F. (1999) Organization and dynamics of the proteolipid complexes formed by Annexin V and lipids in planar supported lipid bilayers. *Biochemistry* **38,** 2779–2786.
22. Chandra, M., da Silva, E. F., Sorenson, M. M., Ferro, J. A., Pearlstone, J. R., Nash, B. E., et al. (1994) The effects of N helix deletion and mutant F29W on the $Ca^{2+}$ binding and functional properties of chicken skeletal muscle troponin. *J. Biol. Chem.* **269,** 14,988–14,894.
23. Moncrieffe, M. C., Venyaminov, S. Y., Miller, T. E., Guzman, G., Potter J. D., and Prendergast F. G. (1999) Optical spectroscopic characterization of single tryptophan mutants of chicken skeletal troponin C: evidence for interdomain interaction. *Biochemistry* **38,** 11,973–11,983.

# 7

# Fluorescence Methods for Measuring Calcium Affinity and Calcium Exchange with Proteins

**J. David Johnson and Svetlana B. Tikunova**

## 1. Introduction

A transient increase in intracellular $Ca^{2+}$ provides the signal for the transient activation of numerous cellular processes including skeletal, cardiac and smooth muscle contraction, neurotransmission, cell proliferation, and division. Nature has designed hundreds of $Ca^{2+}$-binding proteins that sense the "$Ca^{2+}$ signal" and transduce it into cellular action. $Ca^{2+}$-binding proteins are "tuned" to respond to these $Ca^{2+}$ transients by virtue of their specific $Ca^{2+}$ affinities, on-rates, and off-rates. The $Ca^{2+}$-binding parameters of a particular protein dictate the speed, the extent, and the duration of its activation after a $Ca^{2+}$ transient.

In this chapter we describe how $Ca^{2+}$ titrations of fluorescent $Ca^{2+}$-binding proteins are performed, calibrated, and analyzed. We also describe the use of fluorescence stopped-flow methods for determining the rates of $Ca^{2+}$ dissociation and association with proteins.

$Ca^{2+}$-binding to many EF-hand $Ca^{2+}$-binding proteins produce large structural changes and if these structural changes perturb the environment of intrinsic or extrinsic fluorophores, then the $Ca^{2+}$ dependence of these fluorescence changes allow a determination of $Ca^{2+}$ affinity. Examples of this approach are the $Ca^{2+}$-dependent increases in intrinsic tyrosine fluorescence in the C-terminal of TnC and calmodulin (CaM) ***(1–3)*** and the $Ca^{2+}$-dependent increases in extrinsic fluorescent probes in the *N*-terminal of cardiac and skeletal TnC ***(4–6)***. In this chapter, this technique is exemplified by $Ca^{2+}$ titrations of a CaM in which we have mutated F at position 19 to W. This F19W mutant undergoes a large $Ca^{2+}$-dependent increase in TRP fluorescence when $Ca^{2+}$ binds to its *N*-terminal $Ca^{2+}$-binding sites and a large decreases in fluorescence when $Ca^{2+}$

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

is removed from these sites with ethylene glycol-*bis N,N,N',N'*-tetraacetic acid (EGTA). We use fluorescence stopped-flow techniques to measure the rates of the fluorescence changes that occur upon $Ca^{2+}$ association and dissociation from this protein.

The rates of $Ca^{2+}$-induced structural (fluorescence) changes in $Ca^{2+}$-binding proteins may be slower than the actual rates of $Ca^{2+}$ binding and $Ca^{2+}$ dissociation. For this reason, it is necessary to have a method of monitoring $Ca^{2+}$ off rates, which is independent of structural changes in the protein. The fluorescent $Ca^{2+}$ chelator, Quin-2, can be used to measure $Ca^{2+}$ off-rates from native unlabeled proteins ***(7,8)***. In this technique, Quin is rapidly mixed with a $Ca^{2+}$-binding protein with bound $Ca^{2+}$. Quin's fluorescence increases at the rate at which it removes $Ca^{2+}$ from the protein, allowing for a direct determination of $Ca^{2+}$ off-rates. These techniques allow a rapid characterization of the $Ca^{2+}$ binding and exchange properties of any $Ca^{2+}$-binding protein. This information allows for a more accurate prediction of a protein's activation/inactivation profile in response to cellular $Ca^{2+}$ transients.

## 2. Materials

### 2.1. Fluorescence Methods for Determining $Ca^{2+}$ Affinity

1. A scanning fluorescence spectrophotometer.
2. 1 cm path length, four sides polished, 1 mL quartz cuvets.
3. A purified stock of $Ca^{2+}$-binding protein (typically 100–300 μ*M* concentration).
4. A calibrated stock of $CaCl_2$ (generally 0.5 *M*) and a calibrated (*see* **Note 4**) buffer composed of 200 m*M* MOPS, 90 m*M* KCl, and 2 m*M* EGTA at pH 7.0.
5. A $Ca^{2+}$-titration computer printout (as generated by the Robinson and Potter ***[9]***, Fabiato ***[10]***, or Schoenmaker et al. ***[11]*** computer programs) that shows the number of microliters of your $Ca^{2+}$ stock, which should be added to a particular volume of your protein + buffer solution to obtain specific pCas.
6. Fluorescent $Ca^{2+}$ indicators: Quin-2, Fura-2, or Fluo-3.

### 2.2. Methods of Monitoring $Ca^{2+}$ Dissociation and Association Rates from Proteins

1. A fluorescence stopped-flow spectrophotometer with rapid mixing kinetics and a computer for data acquisition and analysis.
2. Purified $Ca^{2+}$-binding proteins, Quin-2, EGTA, and chelex resin.

## 3. Methods

### 3.1. Fluorescence Methods of Monitoring $Ca^{2+}$ Binding to Proteins

**Fig. 1** shows a $Ca^{2+}$ titration of F19W calmodulin as an example of a $Ca^{2+}$ titration of a fluorescent $Ca^{2+}$-binding protein. $Ca^{2+}$ binds half-maximally at pCa 5.4 ($4 \times 10^{-6}$ *M*) and produces a threefold increase in TRP fluorescence

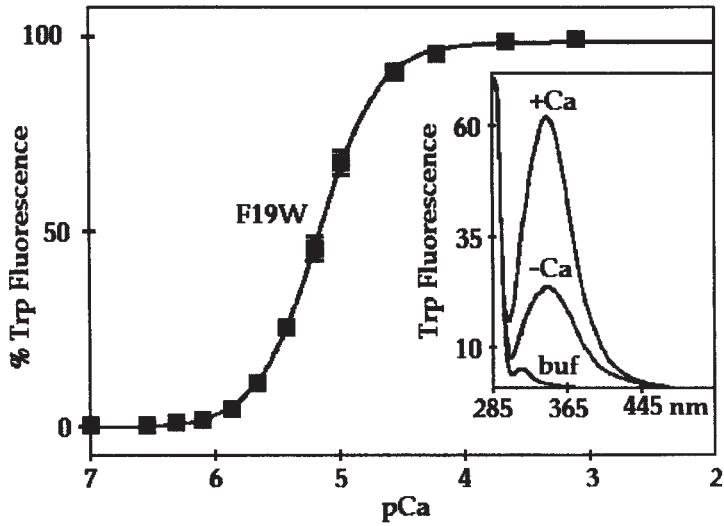


Fig. 1. The $Ca^{2+}$-induced increase in F19W tryptophan fluorescence. $Ca^{2+}$ titrations were conducted as described in the methods section. 100% fluorescence corresponds to a threefold fluorescence increase. Each data point represents the average of three titrations ±S.E. The inset shows fluorescence emission spectra of F19W before (–Ca trace) and after (+Ca trace) the addition of $Ca^{2+}$ (pCa 4.0) and the spectra of the buffer without protein (buf trace).

with a Hill coefficient of 2.0. The inset to **Fig. 1** shows the fluorescence emission spectra of F19W in the absence of $Ca^{2+}$ (–Ca trace) and in the presence of pCa 4.0 (+Ca trace).

To conduct these titrations:

1. Be sure the buffer (200 m*M* MOPS, 90 m*M* KCl, 2 m*M* EGTA at pH. 7.0) and $Ca^{2+}$ stocks have equilibrated to the desired temperature. Add 1 mL of the buffer to a clean cuvet and then add 1 μ*M* of the purified $Ca^{2+}$-binding protein. Place parafilm over the top of the cuvet and mix by inverting three times, being certain that no volume is lost during mixing. Visually inspect the contents of the cuvet and be certain that it has not been contaminated with lint and shows no visible turbidity.
2. Place the cuvet in the fluorimeter and run an emission spectra by setting the excitation to the desired wavelength and scanning the emission wavelength from the excitation wavelength over the wavelength of fluorescence emission. For example, for tryptophan-containing proteins, you can excite at 285 nm and because TRP emissions are generally maximal at 320–360 nm, it is necessary to scan the emission wavelength from 285 to 460 nm, as shown in **Fig. 1**.
3. Fix the emission wavelength to the wavelength where maximum emission was obtained on the emission spectra (335 nm in **Fig. 1**) and record the intensity as a

function of time for approx 1 min. Integrate or average the signal intensity over this time. The fluorescence intensity should be level with time. If not, then it is possible that temperature has not equilibrated because increasing temperature decreases fluorescence. If the intensity increases and decreases, there may be lint in the solution, which increases scattered light (and apparent fluorescence intensity) as it floats through the light path.

4. If the initial intensity is level, begin your $Ca^{2+}$ titrations by adding an aliquot of the $Ca^{2+}$ stock to the cuvet. After each addition of $Ca^{2+}$, mix the solution to assure homogeneity. This is easily accomplished by inserting a pipet tip and drawing 150 μL of solution in and out of the pipet tip five times. Be sure not to push air into the solution, because bubbles forming on the side of the cuvet affect fluorescence.
5. For a complete $Ca^{2+}$ titration, you will want to have ~ 8–10 points ($Ca^{2+}$ concentrations) on the linear portion of the fluorescence increase. If you have an idea of the $K_d$, this can easily be accomplished by hitting 4–5 points 1 Log unit before and 1 Log unit after the $K_d$. $Ca^{2+}$ titrations are generally complete over 2 Log units of [$Ca^{2+}$]. Because $Ca^{2+}$ titrations are often plotted on a Log scale (i.e., pCa), it is important to go high enough (generally 25 times the $K_d$) in [$Ca^{2+}$] to assure leveling.
6. After the last pCa point, run a final emission spectra (+ Ca spectra in **Fig. 1**) over the initial emission spectra. Compare the intensity of the scatter peak (when the emission is at the excitation wavelength, i.e., 285 nm in **Fig. 1**) and its shoulder in the initial and final spectra. If the scatter peaks are similar (as in **Fig. 1**) then the fluorescence changes you are observing are free from interference from scatter or turbidity.
7. Run an emission spectra of the buffer without protein (buf trace in **Fig. 1**). This buffer spectra can also be run first if you know the correct instrument sensitivity for the titration or if you have a fluorimeter with auto scaling.
8. Data analysis. Plot fluorescence intensity as a function of pCa. This can be done in several ways. One method is to plot *F*/*Fo* as a function of pCa. *F* is the fluorescence intensity at each pCa and *Fo* is the initial fluorescence intensity in the absence of added $Ca^{2+}$. Any contribution of buffer should be subtracted before determining *F* and *Fo*. An alternative method is to normalize the maximum fluorescence to 100% and plot% fluorescence increase as a function of pCa (as in **Fig. 1**). Typically, an average of 3–5 titrations are shown with standard error bars. The data is fit with a sigmoidal plot, which allows an accurate determination of $K_d$ and Hill coefficient.

### 3.2. Fluorescence Methods for Monitoring $Ca^{2+}$-Dissociation Rates from $Ca^{2+}$-Binding Proteins

Once you have a $Ca^{2+}$-binding protein which undergoes a large $Ca^{2+}$-dependent change in fluorescence, it is quite easy to determine the rates of $Ca^{2+}$ dissociation using a fluorescence stopped-flow apparatus. Stopped-flow studies with F19W provide a good example of this technique.

1. Before determining the Ca off-rate of a fluorescent $Ca^{2+}$-binding protein it is important to run fluorescent spectra of the protein in the $Ca^{2+}$ free and $Ca^{2+}$ bound state (as in **Fig. 1**). This spectra shows that F19W undergoes a large increase in TRP fluorescence with $Ca^{2+}$-binding. When a $Ca^{2+}$-F19W solution is rapidly mixed with EGTA, its fluorescence should decrease as $Ca^{2+}$ dissociates. Having the fluorescence spectra allows one to choose the correct wavelengths or filter system with which to monitor this fluorescence decrease. For monitoring F19W TRP fluorescence, we use a filter (UG1) that transmits light from 320 to 380 nm. Excitation was at 285 nm as in **Fig. 1**.
2. For measuring $Ca^{2+}$ off-rates using TRP fluorescence: Fill drive-syringe A with 2 mL of 4 μ*M* F19W + 200 μ*M* $CaCl_2$ in a 10 m*M* MOPS, 90 m*M* KCl, pH 7.0 buffer, and the other drive-syringe B with 2 mL of the same buffer + 5 m*M* EGTA. Make sure the temperature in the drive syringes has equilibrated to the desired temperature, in this case 10°C. With our Applied Photophysics stopped-flow, each time the pneumatic drive is activated to initiate a shot, it drives 100 μL of syringe A and 100 μL of syringe B into the mixing chamber and mixing is complete in 1.6 ms. We make three shots to flush the mixing chamber of wash solution (nanopure water) and then begin to collect data. Typically 5–10 shots are collected, averaged, and fit with an exponential equation. **Figure 2** shows the time dependence of the EGTA-induced decrease in F19Ws fluorescence when six shots are averaged. For data analysis, we fit the data after mixing is complete (1.6 ms) with a single- or double-exponential equation using the nonlinear Levenberg-Marquardt algorithm. As shown in **Fig. 2**, F19Ws TRP fluorescence decreases upon $Ca^{2+}$ dissociation as a single exponential (based on $\chi$-squared values and distribution of residuals) at a rate of 310/s.
3. After each series of shots, both syringes and the mixing chamber are flushed with 3–5 mL of nanopure water to remove protein and $CaCl_2$.
4. For the control shots, syringe A has the same F19W + 200 μ*M* $CaCl_2$ solution as aforementioned and syringe B is replaced with a solution of the same buffer and 200 μ*M* $CaCl_2$. 5–10 shots are averaged and this data is overlaid with the EGTA shots above. The control shots give the fluorescence intensity of the $Ca^{2+}$ saturated state and reveal any photobleaching that might be occurring. In our experiments, these control traces were flat lines starting at 4.6 V fluorescence. This showed that no bleaching was occurring over the time range of data acquisition and no correction for bleaching was required. If the control shots exhibit a linear decrease in intensity with time, then photobleaching is occurring and the control shots should be subtracted from the data before fitting. The fact that the control trace starts at 4.6 V also indicates that some of the decreases in fluorescence was lost in the mixing time of the instrument (*see* **Note 5**).

### 3.3. Measurement of $Ca^{2+}$ Dissociation from Proteins Using Quin-2 Fluorescence

The fact that EGTA induces a decreases in F19W fluorescence at 310/s does not prove that this is the rate at which $Ca^{2+}$ is dissociating from this protein. It

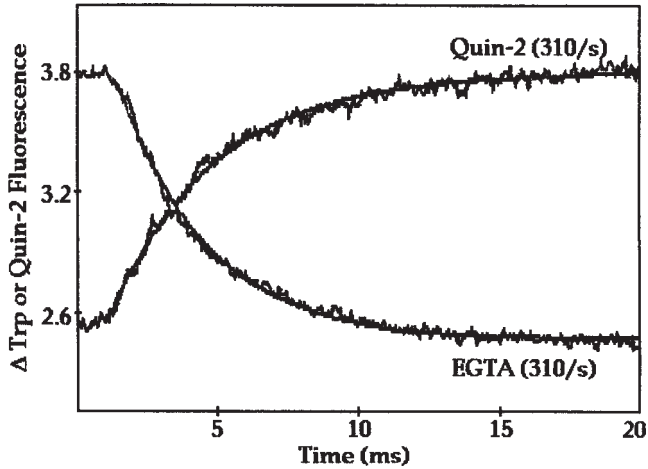


Fig. 2. Rates of $Ca^{2+}$ dissociation from F19W's N-terminal $Ca^{2+}$-binding sites using Quin-2 and TRP fluorescence. Experiments were conducted as described in **Subheading 3**. For Quin experiments, each trace represents an average of 5–7 traces fit with a single exponential (variance $<7.5 \times 10^{-5}$). For the TRP experiments, each trace represents an average of six traces fit with a single exponential (variance $<5 \times 10^{-4}$). All kinetic traces were triggered at zero time, and the first 1.6 ms of premixing are shown (the apparent lag phase).

is possible that $Ca^{2+}$ is dissociating faster than 310/s and that after $Ca^{2+}$ dissociation, a slower conformational change occurs producing the decrease in TRP fluorescence. The fluorescent $Ca^{2+}$ chelator, Quin-2, provides a convenient and accurate method of determining the rate of $Ca^{2+}$ dissociation from proteins that is not dependent upon $Ca^{2+}$-dependent structural changes (*see* **Notes 6** and **7**).

1. For measuring $Ca^{2+}$ off-rates using Quin-2 fluorescence: Fill drive syringe A with 2 mL of 8 μ*M* F19W + 60 μ*M* $CaCl_2$ in a 10 m*M* MOPS, 90 m*M* KCl, pH 7.0 buffer and the other drive syringe B with 2 mL of the same buffer + 150 μ*M* Quin.
2. Shoot three shots to flush wash solution out of the mixing chamber.
3. Shoot and average 5–10 shots of the aforementioned reaction to obtain the data shown in **Fig. 2** (Quin-2 trace). This data shows that Quin fluorescence increases as a single exponential at 310/s. This is the same rate as F19W TRP fluorescence decreases upon mixing with EGTA. Thus, both methods report the same rate for $Ca^{2+}$ dissociation and the changes in F19W TRP fluorescence are accurately reporting the rate of $Ca^{2+}$ dissociation from this protein.
4. Shoot a series of control shots where 10, 20, 30, and 60 μ*M* $Ca^{2+}$ is reacted with 150 m*M* Quin. Average 5 shots at each [$Ca^{2+}$]. These can be used to calculate the moles of $Ca^{2+}$ dissociating from the protein and to correct for any photobleaching (*see* **Note 8**).

### 3.4. Measurements of $Ca^{2+}$ On-Rates

$Ca^{2+}$ on-rates can be measured in proteins like F19W by observing the rate of the $Ca^{2+}$-induced increase in TRP fluorescence as a function of increasing [$Ca^{2+}$].

1. In order to measure $Ca^{2+}$ on rates, the protein should be $Ca^{2+}$ free. If it is not, this is achieved by incubating approx 30 μ*M* of protein with chelex resin (generally 1 mL of resin for each 4 mL of protein) for approx 4 h in a plastic test tube. The solution should be shaken, not stirred during the incubation. Stir bars can fragment the resin. After incubation, the resin is allowed to settle to the bottom or is pelleted by low-speed centrifugation and the supernatant decanted. The chelexed protein can then be diluted into chelexed buffer and tested for $Ca^{2+}$ occupancy. This is done by diluting 2 μ*M* of the protein into 1 mL of a chelexed buffer (10 m*M* MOPS, 90 m*M* KCl, pH 7.0) and running a fluorescence spectra, as shown in **Fig. 1**. After taking the initial spectra, add 100 μ*M* $Ca^{2+}$ to the protein and run the spectra of the protein in the $Ca^{2+}$ saturated state. Finally, add 2 m*M* EGTA to this 1-mL solution to produce the $Ca^{2+}$ free state and run a spectra. By comparing the fluorescence intensity in the original state, the + $Ca^{2+}$ state and the +EGTA state, the percent saturation of the chelexed protein can be easily determined. If the protein is found to be $Ca^{2+}$ free, then it can be used for determining $Ca^{2+}$ on rates (*see* **Note 10**).
2. Fill syringe A with 2 μ*M* protein (F19W) in 10 m*M* MOPS, 90 m*M* KCl at pH 7.0. Fill syringe B with the same chelexed (if required) buffer.
3. After three shots to clear the mixing chamber of wash solution, average 5–7 shots. Because we are introducing little or no $Ca^{2+}$ to protein in these control shots, there should be little time dependent increase in F19W fluorescence.
4. Keeping syringe A the same, now add increasing amounts of $Ca^{2+}$ (4, 6, 8, 10, 15, and 20 μ*M*) to syringe B and average 5–7 shots at each [$Ca^{2+}$]. The [$Ca^{2+}$] points to be used can be determined from the $Ca^{2+}$ titration of the protein (*see* **Fig. 1**). The rate of the increase in F19W TRP fluorescence should increase as a function of increasing [$Ca^{2+}$] as shown in **Fig. 3** inset. F19W TRP fluorescence increases at a rate of 495/s for 2 μ*M* $Ca^{2+}$ (after 1:1 dilution of 4 μ*M* $Ca^{2+}$), at 658/s for 4 μ*M* $Ca^{2+}$ and at 1138/s for 10 μ*M* $Ca^{2+}$. **Figure 3** shows a plot of the rate ($K_{obs}$) of the $Ca^{2+}$-induced increase in F19W fluorescence as a function of increasing [Ca]. This linear plot exhibits a slope which is equal to the $Ca^{2+}$ association rate $8 \times 10^7$ M/s and an intercept on the *y*-axis (approx 320/s) should equal to the $Ca^{2+}$ off-rate. For further verification of this on rate *see* **Note 11**.

## 4. Notes

### 4.1. Fluorescence Methods for Determining $Ca^{2+}$ Affinity

Two major advantages of using fluorescence changes to follow $Ca^{2+}$-binding are: The ease and reproducibility of the $Ca^{2+}$ titrations; (each titration takes approx 30 min and error bars of <5% are easily obtained) and the low concentrations of protein required (typically 0.1–1 μ*M*). In addition, most purified

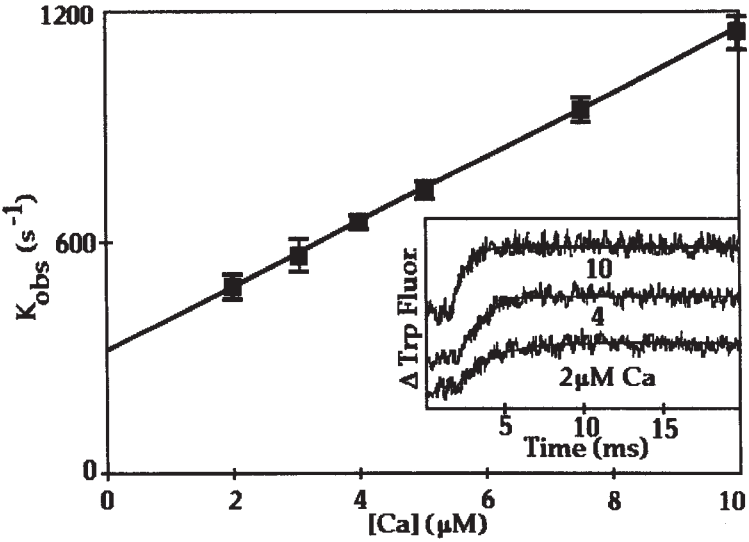


Fig. 3. Rates of $Ca^{2+}$ association with F19W as a function of increasing [$Ca^{2+}$]. A plot of the observed rate of increase in F19W fluorescence (Kobs) vs the free [$Ca^{2+}$] in the mixing chamber before binding is shown. The inset shows the rate of the increase in F19W TRP fluorescence when F19W is rapidly mixed increasing concentrations of $Ca^{2+}$ as described in the methods section at 10°C. Each curve represents an average of 5–7 traces fit with a single exponential (variance $< 9 \times 10^{-4}$). Each point represents the average of three experiments ± S. E.

$Ca^{2+}$-binding proteins are incredibly stable and can be stored at –20°C for years, without degrading. There are several other factors which can interfere with accurate $Ca^{2+}$ titrations.

1. Inner filter effects — If titrations are done with very high concentrations of protein or fluorophore this can result in the protein absorbing excitation light near the surface or face of the cuvet resulting in less excitation and emission from the center of the cuvet. Although these inner-filter effects are not generally a problem when 10 μ*M* or less protein or fluorescent probe are used, they can produce artifacts at higher concentrations. Inner-filter effects depend not only on the extinction coefficient and concentration of the fluorophore, but also the path length of the cuvet and on the optics of the fluorimeter. Generally, it is advisable that concentrations be used so that the absorption of the fluorophore is less than 0.2 OD.
2. Dilution of the fluorophore decreases fluorescence — The fluorescence intensity of any fluorophore will decrease as the sample is diluted. Because fluorescence changes are generally large, dilution is not generally a problem. For a $Ca^{2+}$ titration, we generally add less than 30 μL of total $Ca^{2+}$ stock to a 1-mL cuvet and this would dilute the fluorescence intensity by only 3% from its original value. Dilu-

tion can be minimized by using more concentrated $Ca^{2+}$ stocks. If necessary, the fluorescence intensity can be corrected for dilution at each point in the titration by multiplying the observed fluorescence intensity after the addition by the dilution factor (the ratio of the volume of the solution after the addition to the volume of the solution before the addition).

3. Always run spectra — Even though it takes a few additional minutes, it is essential to run an emission (or excitation spectra) of your solution both before and after the titration. This assures that the fluorescence change you are following results from a change in the fluorescence of the fluorophore and is not a result of changes in light scattering. These spectra allow one to determine the wavelength at which the largest fluorescence change occurs and this can be used to increase signal to noise in the titration. Although it is generally best to do the titration using the emission wavelength where the greatest fluorescence change occurs, in more turbid solutions, scattering can be a problem. When this is the case, the contribution of scatter to emission can be reduced by selecting an emission wavelength that is further removed from the excitation wavelength and where the fluorophore still undergoes a large fluorescence change with $[Ca^{2+}]$.
4. Calibration of $CaCl_2$ stocks and EGTA buffers — The most critical points for accurate $Ca^{2+}$ titrations is the precise control of free $[Ca^{2+}]$ using $Ca^{2+}$ chelators like EGTA or ethylenediaminetetracetic acid (EDTA). It is important to use sufficient $Ca^{2+}$ buffer (EGTA or EDTA) so that the small amount of endogenous $Ca^{2+}$ bound to the protein (generally 1 μ*M* or less) being titrated does not affect free $[Ca^{2+}]$. We typically conduct titrations in the presence of 2 m*M* EGTA and use 200 m*M* MOPS (or HEPES) at pH 7.0 to prevent decreases in pH as $Ca^{2+}$ binds and releases protons from EGTA. Other buffers can be used as long as they do not contribute or bind $Ca^{2+}$. Any desired ionic strength can be used as long as its affect on EGTAs affinity for $Ca^{2+}$ are considered by the pCa computer program. Several computer programs are available for determining the free $Ca^{2+}$, as a function of total added $Ca^{2+}$, for a given [EGTA] or [EDTA] at various temperatures, ionic strengths and pH (*see* **Notes 9** and **10**). Obviously, the stock $[CaCl_2]$ and [EGTA] must be accurately made and this is facilitated by heating the powders of both to 80–90°C for 10 min to remove water before weighing.

   The $CaCl_2$ and EGTA stocks and the ability to accurately vary free $Ca^{2+}$ as a function of added $Ca^{2+}$ should always be tested and calibrated by conducting $Ca^{2+}$ titrations of a fluorescent $Ca^{2+}$ indicator like Quin-2, Fura-2, or Fluo-3. Because each of these $Ca^{2+}$ indicators bind one mole of $Ca^{2+}$ per mole of indicator with known $K_d$s, if the fit of the $Ca^{2+}$ dependent change in fluorescence gives the appropriate $K_d$ and Hill coefficient (1.0) you can be confident that you are accurately controlling free $Ca^{2+}$. **Figure 4** shows typical $Ca^{2+}$ dependent increases in Quin-2, Fura-2, and Fluo-3 fluorescence. Half-maximal increases occurred at pCa 7.21 (62 n*M*) for Quin-2, pCa 6.86 (138 n*M*) for Fura-2, and pCa 6.36 (430 n*M*) for Fluo-3. All of the indicators exhibited Hill coefficients of 1.03 ± 0.04. These $K_d$s are essentially identical to the $K_d$s for Quin-2 (60 n*M*), Fura-2 (145 n*M*) and Fluo-3 (390 n*M*), reported by their supplier in the Molecular Probes

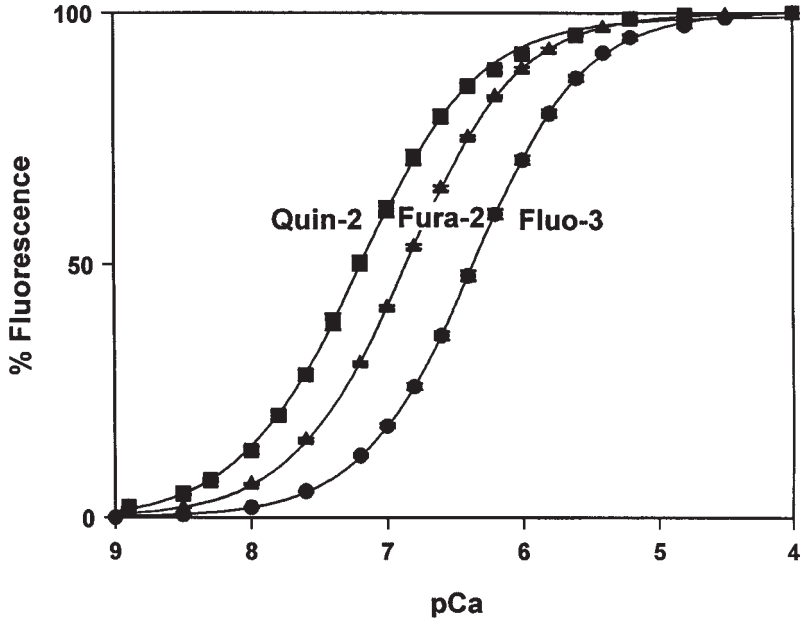


Fig. 4. $Ca^{2+}$ titrations of Quin-2, Fura-2, and Fluo-3. Increasing concentrations of $Ca^{2+}$ were added to 1 μ*M* of each indicator in 1 mL of buffer (200 m*M* MOPS, 90 m*M* KCl, 2 m*M* EGTA) at 22°C. Excitation was at 330, 340, and 490 nm and emission was at 495, 510 and 525 nm for Quin-2, Fura-2, and Fluo-3, respectively. 100% fluorescence increase corresponds to 7.8-, 2.6-, and 52-fold increases for Quin-2, Fura-2, or Fluo-3, respectively. Each data point represents an average of 3 titrations ± S. E.

handbook ***(12)***. Thus, $Ca^{2+}$ titrations of any of these indicators provides an easy means of verifying the accuracy of your $CaCl_2$ and EGTA stocks.

Errors in either $K_d$ or Hill coefficient can indicate inappropriate control of pCa. If this occurs, then the concentrations of the EGTA or $CaCl_2$ stocks must be off and should be adjusted till accurate $Ca^{2+}$ dependent increases in indicator fluorescence are obtained. The concentration of the $CaCl_2$ stock (generally 0.5 *M*) can also be verified by atomic absorption spectroscopy. If the $CaCl_2$ stock is correct, errors in the [EGTA] are probably responsible for deviant $Ca^{2+}$ titration. After your $CaCl_2$ and EGTA buffers have been made and calibrated, they can be broken into aliquots and stored frozen, in plastic, for years.

### 4.2. Fluorescence Methods to Measure $Ca^{2+}$ Dissociation Rates from Proteins

5. For very rapid kinetic reactions, some of the fluorescence change may be lost during in the mixing time of the instrument. For example in **Fig. 2**, F19W + $Ca^{2+}$ had a fluorescence intensity of 4.6 V (control shot) yet the observable decrease in

its fluorescence began at 3.8 V. The exponential fit of this data extrapolates back to 4.6 V and this is consistent with the fact that for a reaction occurring at 310/s approx 36% of the fluorescence change (0.8 V) would occur in the 1.6 ms mixing time. The extrapolated exponential fit of this data corrects for this loss in signal or, if desired, the reaction temperature can be reduce to slow the off rate so that more of the change can be observed.

6. Quin can be used to measure $Ca^{2+}$ dissociation rates from essentially any $Ca^{2+}$-binding protein and this method does not require that the $Ca^{2+}$-binding protein has a fluorescent label. Repeating our Quin shots with wtCaM (over 0–20 ms) showed that $Ca^{2+}$ dissociates from the N-terminal sites of wtCaM nearly twice as fast as from F19W. This suggests that the F mutation is slightly slowing the $Ca^{2+}$ off-rate. In addition, if these Quin experiments (*see* **Fig. 2**) are conducted over longer times (0–2 s) you can observe $Ca^{2+}$ dissociating from the slower C-terminal sites of F19W CaM at 3.5/s. Thus, Quin is extremely useful for verifying the effect of any mutation, drug, peptide, or protein binding on $Ca^{2+}$ off-rates from essentially any purified $Ca^{2+}$-binding protein ***(8,13,14)***. The exception to this is that if $Ca^{2+}$ affinity is reduced to much, the rates of $Ca^{2+}$ dissociation may become too rapid to observe.
7. Quin is well suited for these type of measurements because of three factors:
   a. It has a diffusion limited $Ca^{2+}$ on rate;
   b. It has a high affinity for $Ca^{2+}$ and is an effective chelator;
   c. It undergoes a large fluorescence increases upon binding $Ca^{2+}$.

   The fact that Quin has a diffusion limited $Ca^{2+}$ on-rate means that as soon as the $Ca^{2+}$ dissociates from the protein it will bind Quin and any $Ca^{2+}$ that is not bound to the protein will bind to Quin during the mixing time of the instrument. These effects are facilitated by keeping the [Quin] in excess of the [$Ca^{2+}$] and [$Ca^{2+}$-binding protein]. Quin's dissociation of $Ca^{2+}$ from proteins is a second-order reaction dependent on the concentration of Quin and $Ca^{2+}$ -protein. It is, therefore, important to use enough Quin to chelate $Ca^{2+}$ at its true $Ca^{2+}$ dissociation rate (make the reaction pseudo first order). In the above experiments we mixed 8 μ*M* of $Ca^{2+}$-loaded protein with 150 μ*M* Quin. In subsequent experiments, we verified that increasing [Quin] (to 200 μ*M*) and decreasing [$Ca^{2+}$] (from 60 to 30 μ*M*) did not increase the $Ca^{2+}$ off-rate. For higher affinity $Ca^{2+}$-binding proteins, the [$Ca^{2+}$] concentration can be reduced and the Quin concentration can be increased to assure complete and first-order removal of $Ca^{2+}$. Generally it is better to decrease the [$Ca^{2+}$] because if [Quin] is increased much above 200 μ*M* (in our instrument), inner-filter effects may produce reductions in the signal to noise because of Quin's absorption of exciting light. Typically, we do a Quin experiment like the one described at several different [Ca] and [Quin] to verify that the off-rate is not altered.
8. Another advantage of the Quin technique is that it can be used to estimate the number of moles of $Ca^{2+}$ that are dissociating from the protein. This is accomplished by rapidly mixing 150 μ*M* Quin with increasing concentration of $Ca^{2+}$ (10, 20, 30, and 60 μ*M*) and determining the relationship between Quin fluores-

cence intensity and bound $Ca^{2+}$. These calibration studies should be done directly after the $Ca^{2+}$ off-rate experiments, using the same instrument conditions, to reduces changes in intensity or amplitude produced by time dependent changes in lamp output. By comparing the amplitude of the increase in Quin fluorescence in the $Ca^{2+}$ off-rate experiments to the increase in Quin fluorescence when it is reacted with a specific [Ca], we calculate that two moles of $Ca^{2+}$ were dissociated per mole of F19W in **Fig. 2**. Thus, Quin allows a rapid determination of not only the $Ca^{2+}$ off-rate, but also the number of sites from which $Ca^{2+}$ dissociated.

9. Quin will report the $Ca^{2+}$ dissociation rate from any contaminating chelator that is with the protein. Thus, it is important to thoroughly dialyze your purified protein to remove chelator. If the protein is contaminated with EGTA or EDTA you will see $Ca^{2+}$ dissociation at 0.55/s and 0.7/s, respectively, at 10°C, because of $Ca^{2+}$ dissociation from chelators ***(15)***.

### 4.3. Measurement of $Ca^{2+}$ On-Rates

10. While chelex can generally be used to remove $Ca^{2+}$ from the protein, for high-affinity $Ca^{2+}$-binding proteins some residual $Ca^{2+}$ may remain. In these cases, small amounts of EGTA can be used to remove the remaining $Ca^{2+}$. It should be noted that EGTA can affect the measured on rate for proteins that have slower on rates. This is because EGTA can bind some of the $Ca^{2+}$ which is being reacted with the protein at a rate of $1.3 \times 10^6$ *M*/s (*see* **ref. *14***) and at higher [EGTA] this effect is substantial .
11. A quick approximation of the $Ca^{2+}$ on-rate can be obtained before shooting all of the [$Ca^{2+}$] shown in **Fig. 3** by simply mixing a 10- to 20-fold molar excess of $Ca^{2+}$ with the protein and determining the rate of increase in protein fluorescence. When $Ca^{2+}$ is in sufficient excess of protein, the reaction approaches pseudofirst-order and the on-rate can be approximated by the ratio of the observed on rate (1138/s) to the concentration of $Ca^{2+}$ reacted with the protein (10 μ*M*), yielding an on-rate of $1.1 \times 10^8$ *M*/s. Further verification of the $Ca^{2+}$ on-rate can be achieved by confirming that the off-rate (approx 320/s) determined as the intercept on the *y*-axis in the linear plot of **Fig. 3** inset, is similar to the actual measured off-rate (310/s in **Fig. 2**).

    Because the $K_d$ is the ratio of the off rate to the on rate, the on rate can also be calculated from $K_{on} = K_{off}/K_d$. For F19W the $K_{off}$ is 310/s and the $K_d$ (from **Fig. 1**) is $4 \times 10^{-6}$ *M*, indicating a $K_{on}$ of $8 \times 10^7$ *M*/s, identical to the actual measured $Ca^{2+}$ on-rate. Thus, by using Quin and TRP fluorescence we can completely characterize the $Ca^{2+}$ affinity and the $Ca^{2+}$ exchange rates of a $Ca^{2+}$-binding protein.

### 4.4. Other Methods for Determining $Ca^{2+}$ Binding and $Ca^{2+}$ Exchange Rates with Proteins

12. $Ca^{2+}$-binding to proteins can also be completely analyzed in terms of affinity, number of sites, and cooperativity of sites by using $^{45}Ca^{2+}$ and equilibrium or flow dialysis (as discussed by M. Yazawa in Chapter 1) and by the use of $Ca^{2+}$

chelators that undergo changes in absorption (or fluorescence) upon $Ca^{2+}$-binding (as discussed by S. Linse in Chapter 2).

13. The rates of $Ca^{2+}$ dissociation from proteins can also be determined using the luminescent trivalent cation terbium ($Tb^{3+}$). $Tb^{3+}$ undergoes a large increase in fluorescence (an phosphorescence) upon binding to $Ca^{2+}$-binding proteins. When $Tb^{3+}$ is reacted with proteins containing bound $Ca^{2+}$ (or $Mg^{2+}$), its luminescence increases at the rate of $Ca^{2+}$ (or $Mg^{2+}$) dissociation. We have used $Tb^{3+}$ fluorescence and stopped-flow methodology to determine the rates of $Ca^{2+}$ and $Mg^{2+}$ dissociation from parvalbumin and EGTA ***(16)***. $Ca^{2+}$ exchange rates can also be determined by NMR spectroscopy (*see* T. Drakenberg, Chapter 18).
14. We have recently introduced a method that allows the generation of $Ca^{2+}$ transients of various amplitudes and duration in a stopped-flow apparatus. This method is based on the fact that EGTA and Mg-EDTA have a slow $Ca^{2+}$ on-rate and when $Ca^{2+}$ is rapidly mixed with a solution containing EGTA (or Mg-EDTA), [$Ca^{2+}$] transiently rises until it is bound by chelator. This has allowed us to produce "artificial" $Ca^{2+}$ transients which vary in duration from 0.1 to 50 ms and to observe the transient activation of various $Ca^{2+}$-binding proteins and $Ca^{2+}$ dependent enzymes in response to these transients ***(15)***. Thus, using stopped-flow techniques similar to those discussed above it is possible to produce $Ca^{2+}$ transients and follow the response of $Ca^{2+}$-binding proteins to these transients in a stopped-flow apparatus.

## Acknowledgments

This work was supported by a grant from the National Institutes of Health (DK33727).

## References

1. Johnson, J. D. and Potter J. D. (1978) Detection of two classes of calcium binding sites in troponin C with circular dichroism and tyrosine fluorescence. *J. Biol. Chem.* **253,** 3775–3777.
2. Dedman, J. R., Potter, J. D., Jackson, R. L., Johnson, J. D., and Means, A. R. (1977) Physicochemical properties of rat testis calcium-dependent regulator protein of cyclic nucleotide phosphodiesterase. *J. Biol. Chem.* **252,** 8415–8422.
3. George, S. E., Su, Z., Fan, D., and Means, A. R. (1993) Calmodulin-cardiac troponin C chimeras. Effects of domain exchange on calcium binding and enzyme activatin. *J. Biol. Chem.* **268,** 25,213–25,220.
4. Pearlstone, J. R., Borgford, T., Chandra, M., Oikawa, K., Kay, C. M., Herzberg, O., Moult, J., Herklotz, A., Reinach, R. C., and Smillie, L. B. (1992) Construction and characterization of a spectral probe mutant of troponin C: application to analyses of mutants with increased calcium affinity. *Biochemistry* **31,** 6545–6553.
5. Johnson, J. D., Collins, J. H., and Potter, J. D. (1978) Dansylaziridine-labeled Troponin C: a fluorescent probe of calcium binding to the calcium specific regulatory sites. *J. Biol. Chem.* **253,** 6451–6458.

6. Johnson, J. D., Collins, J. H., Robertson, S. P., and Potter, J. D. (1980) A fluorescent probe study of calcium binding to the calcium specific sites of cardiac troponin and troponin C. *J. Biol. Chem.* **255,** 9635–9640.
7. Bayley, P., Ahlstrom, P., Martin, S. R., and Forsen, S. (1984) The kinetics of calcium binding to calmodulin: Quin 2 and ANS stopped-flow fluorescence studies. *Biochem. Biophys. Res. Commun.* **120,** 185–191.
8. Martin, S. R., Maune, J. F., Beckingham, K., and Bayley, P. (1992) Stopped-flow studies of calcium dissociation from calcium-binding -site mutants of Drosophila melanogaster calmodulin. *Eur. J. Biochem.* **205,** 1107–1114.
9. Robertson, S. and Potter, J. D. (1984) The regulation of free calcium ion concentration by metal chelators. *Methods Pharmacol.* **5,** 63–75.
10. Fabiato, A. (1988) Computer programs for calculating total from specificed free or free from specifed total ionic concentrations in aqueous solutions containing multiple metal ligands. *Methods Enzymol.* **157,** 378–417.
11. Schoenmakers, T. J., Visser, G. J., Flik, G., and Theuvenet, A. P. (1992) Chelator: an improved method for computing metal ion concentrations in physiological solutions. *Biotechniques* **6,** 870–874.
12. Haugland, R. P. (1996) *Molecular Probes Handbook of Fluorescent Probes and Research Chemicals*, (Spence, M. T. Z., ed.), 6th ed., Molecular Probes, Inc., Europe, UK, p. 505.
13. Johnson, J. D., Snyder, C., Walsh, M. P., and Flynn, M. (1996) Effects of myosin light chain kinase and peptides on calcium exchange with the N- and C-terminal calcium binding sites of calmodulin. *J. Biol. Chem.* **271,** 761–767.
14. Brown, S. E., Martin, S. R., and Bayley, P. M. (1997) Kinetic control of the dissociation pathway of calmodulin-peptide complexes. *J. Biol. Chem.* **272,** 3389–3397.
15. Davis, J. P., Tikunova, S. B., Walsh, M. P., and Johnson, J. D. (1999) Characterizing the response of calcium signal transducers to generated calcium transients. *Biochemistry* **38,** 4235–4244.
16. Johnson, J. D., Jiang, Y. D., and Rall, J. A. (1999) Intracellular EDTA mimics parvalbumin in the promotion of skeletal muscle relaxation. *Biophys. J.* **76,** 1514–1522.

# 8

# Surface Plasmon Resonance of Calcium-Binding Proteins

**Karin Julenius**

## 1. Introduction

Surface plasmon resonance (SPR) is an optical phenomenon used in certain commercial instruments to measure the kinetics of interaction between macromolecules. One of the interacting counterparts is immobilized on a sensor chip surface, whereas the other is present in the solvent above the surface. The SPR response is correlated to changes in refractive index at the sensor chip surface caused by concentration changes, e.g., when the analyte binds to the immobilized ligand. The SPR signal is monitored continuously which makes it possible to measure both association and dissociation rate constants ($k_{on}$ and $k_{off}$). Once these are established, the equilibrium binding constant ($k_a$) can be calculated.

$$K_a = k_{on} / k_{off}$$

At an interface between two transparent media of different refractive indexes (e.g., glass and water), light coming from the side of higher refractive index is partly reflected and partly refracted. Above a certain critical angle of incidence, no light is refracted and total reflection is observed. At the same time, an electromagnetic field component the so-called evanescent wave penetrates a short distance (of the order of one wavelength) into the medium of lower refractive index. If the interface between the media is coated with a thin layer of metal, and the light is monochromatic and p-polarized (i.e., the electric vector component is parallel to the plane of incidence), the intensity of the reflected light is markedly reduced at a specific incident angle (*see* **Fig. 1**). This phenomenon is called SPR ***(1)***. The SPR angle depends on the properties of the metal film (type of metal, optical constants, thickness, uniformity, and so on), the wave-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

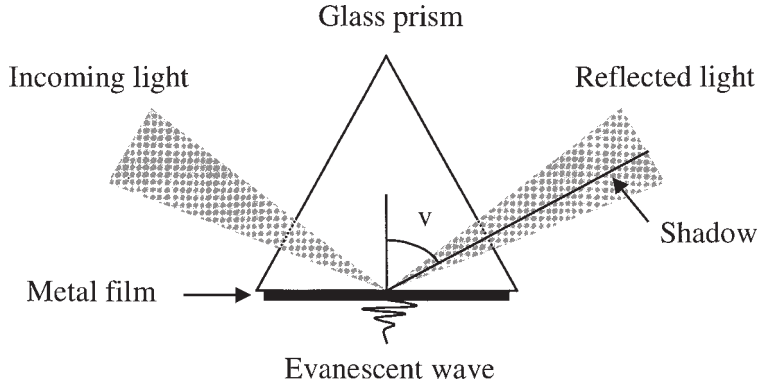


Fig. 1. Under conditions of total internal reflection at a metal-coated interface, an evanescent wave propagates into the medium of lower refractive index. Surface plasmon resonance is observed as a sharp dip in reflected intensity at an angle which depends on the refractive index of the medium on the nonilluminated side of the surface.

length of the incident light, and the refractive index of the media on either side of the metal film. If the metal film properties, wavelength, and refractive index of the denser medium is kept constant, the SPR signal can be used to probe the refractive index of the medium into which the evanescent wave propagates (the nonilluminated side of the surface), as is done in the commercial applications.

The SPR instrument utilizes a sensor chip, which consists of a glass slide with a thin layer of metal on one side. The metal film is covered with a matrix, to which the macromolecules are immobilized. The chip can be replaced, making it possible to use different kind of matrices depending on the nature of the molecule to be immobilized (*see* **Note 1**). The need for substantial changes in refractive index upon binding limits the use to macromolecules; i.e., it is not possible to measure calcium-binding directly. But it is possible to study the calcium dependence of interactions between macromolecules by performing experiments in the absence and presence of calcium. This chapter will describe the testing of calcium-dependence of a protein–protein association reaction using standard coupling methods on a carboxymethylated dextran matrix chip. It will also describe a quantitative kinetic experiment.

## 2. Materials

1. SPR instrument. *See* **Note 2** for information about commercial instruments.
2. Sensor chip with carboxymethylated dextran matrix surface. These are bought from the manufacturer of the SPR instrument. BIAcore calls this chip CM5.
3. Normalization solution: glycerol dissolved in distilled water is needed if the sensor chip is new. The glycerol concentration needed is instrument-dependent, see your instrument manual.

4. Flow buffers (store at 4°C, stable for 2 wk):
   a. 10 m*M* HEPES, 3.4 m*M* ethylenediaminetetracetic acid (EDTA), 0.15 *M* NaCl, approx 0.005% Tween-20 (a surfactant; *see* **Note 4**), approx 0.02% $NaN_3$ (toxic), pH 7.4.
   b. 10 m*M* HEPES, 2 m*M* $CaCl_2$, 0.15 *M* NaCl, approx 0.005% Tween-20, approx 0.02% $NaN_3$ (toxic), pH 7.4.
5. Immobilization buffers (store at 4°C, stability depending on pH): 10 m*M* sodium acetate, adjust pH with 5 *M* acetic acid. Make 3–4 with different pH, slightly under the isoelectic point of the protein to be immobilized (e.g., pH 3.5, 4.0, and 4.5 for a protein with isoelectric point approx 4.5).
6. EDTA and $CaCl_2$ solutions:
   a. 10 m*M* EDTA, pH 8.0.
   b. 0.5 *M* EDTA, pH 7.4.
   c. 0.5 *M* $CaCl_2$.
7. 0.1 *M* HCl.
8. Sterile, disposable filters, 0.22 μm (e.g., Corning):
   a. Syringe tip filters ∅ approx 25 mm and 5 mL sterile, disposable syringes to go with them.
   b. Bottle top filters (either 33- or 45-mm neck size, depending on what kind of glass bottles you readily have; *see* **Note 5**).
9. Coupling reagents:
   a. 0.1 *M N*-hydroxysuccinimide (NHS).
   b. 0.4 *M N*-ethyl-N'-(dimethylaminopropyl)carbodiimide (EDC).
   c. 1 *M* ethanolamine hydrochloride.

## 3. Methods

### 3.1. Preparation

1. Filter all solutions. The amounts needed are such that the flow buffers should be filtered using the bottle top filters and all other solutions should be filtered using syringe tip filters. Do not forget to filter the protein stock solutions. Solutions older than 1 wk should be filtered again before use.
2. Degas the flow buffers under vacuum for approx 5 min.
3. Start the SPR instrument and insert the sensor chip (see the instrument manual for a detailed description of all instrument-related instructions).
4. Place the pump inlet tubing in the calcium-buffer flask. Place a beaker at the waste outlet. Initiate the flow system with the new buffer.
5. Set the temperature to the desired value. Wait until it is equilibrated.
6. If the chip is new, it should be normalized using a glycerol solution.

### 3.2. Immobilization

1. Set the flow rate to 5 μL/min.
2. Wait for equilibrium baseline and note the SPR signal. This is our first reference point.

3. Mix 50 μL of NHS solution with 50 μL of EDC solution. Inject 40 μL of the surface activation mixture into one of the flow cells on the sensor chip (*see* **Note 6**). With an automated instrument, both mixing and injecting can be done automatically.
4. Note the SPR signal after the surface activation is finished. This is our second reference point.
5. Make 100 μL of an immobilization mixture using the protein to be immobilized (*see* **Note 7**). Use one of the immobilization buffers, 10–100 μg/mL of protein and 1 m*M* $CaCl_2$ if the protein binds calcium. Inject 45 μL of the immobilization mixture.
6. Note the SPR signal after the immobilization is finished. Compare it to the second reference point. The difference is an estimate of the protein now immobilized to the surface. 1000 response units (RU) given by a BIAcore instrument equal a protein surface concentration of about 1 ng/mm$^2$. An absolute minimum is that it should be significantly larger than the instrument noise. For a very small peptide, this may be enough, but for a medium-sized protein it should be larger. Typical responses for surface binding of proteins are of the order of 100–20,000 RU (*see* **Note 8**). If it seems like the immobilization has not worked properly, it is possible to try again, as long as the surface has not been deactivated. Go back to **step 4** and change one or more of the conditions (change the pH, the protein concentration, the injected volume, the flow rate, replace $CaCl_2$ by EDTA, use neither $CaCl_2$ nor EDTA, and so on).
3. Deactivate the surface by injecting 20 μL ethanolamine hydrochloride.
4. Free the surface of any noncovalently attached protein molecules (regeneration) by injecting 8 μL of a regenerator. For a calcium-dependent association (*see* **Subheading 3.3.**), one may use 10 m*M* EDTA. Otherwise, 0.1 *M* HCl is the standard regenerator. Other examples of regenerators are found in the manual of the instrument.
5. Note the SPR signal after the regeneration. This should be compared to the first reference point and the difference is the true amount of immobilized protein.

### 3.3. Qualitative Experiment: Is the Interaction Calcium-Dependent?

1. Mix the analyte stock with calcium buffer to obtain three different concentrations between 1 n*M* and 1000 n*M*. Use lower concentrations for large proteins and higher for small proteins. Each sample should be 150 μL. Make the same concentrations of samples dissolved in EDTA buffer.
2. *See* to that the pump inlet tubing is in the calcium-buffer flask and do not forget to initiate the flow system if the buffer has been changed. Set the flow rate to 5 μL/min.
3. Start recording the SPR signal (start a sensorgram).
4. Inject 75 μL of one of the calcium samples.
5. After the association phase is over, keep on recording the signal for approx 30 min.
6. Inject 8 μL regenerator.
7. Stop recording the signal. A typical sensorgram is shown in **Fig. 2**.

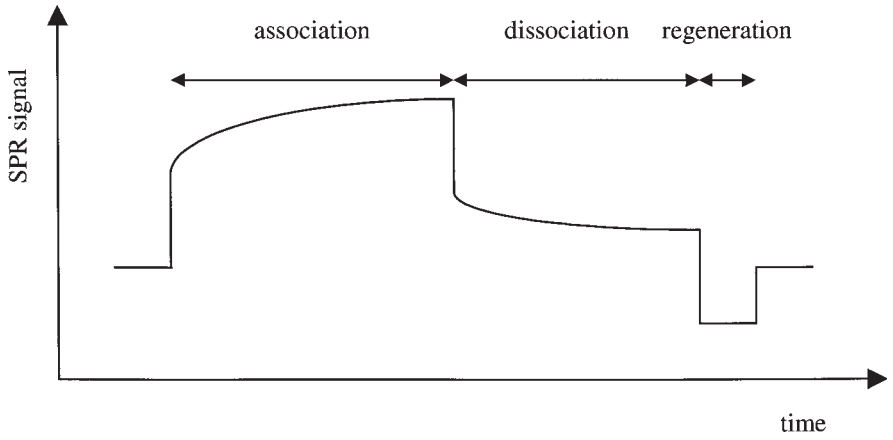


Fig. 2. A typical sensorgram. The sharp rise in signal during the first couple of seconds of the association phase is an effect of the bulk concentration change, as is the sharp decrease in signal at the beginning of the dissociation phase.

8. Repeat from **step 3** using the next calcium sample.
9. When the three calcium experiments have been performed, repeat **steps 3–7** using buffer only as a control. This sensorgram should look like **Fig. 3**.
10. Place the pump inlet tubing in the EDTA-buffer flask and initiate the flow system with the new buffer.
11. Repeat between **steps 3** and **9** for the three EDTA samples and a buffer control.
12. If calcium is absolutely essential for the interaction, the association and dissociation phases of the calcium samples will look like the one in **Fig. 2** and the sensorgrams of the EDTA samples will look like **Fig. 3** (*see* **Note 10**). If the EDTA samples interact with the surface, but weaker than in the calcium case, the association will be slower and/or the dissociation faster. *See* below for quantitative measurement of kinetics.

### 3.4. Quantitative Kinetic Experiment

1. Use a calcium or EDTA buffer depending on what you want to measure. Do not forget to initiate the flow system if you have changed the flow buffer. Set the flow rate to 5 μL/min.
2. Make a few test runs to determine a good concentration interval (follow **steps 3–7**).
3. Make six samples with different analyte concentration within the appropriate concentration range dissolved in the flow buffer. The dispersion between the concentrations should be approximately a factor of 2 (e.g., 1, 2, 5, 10, 20, and 50 n*M*). Each sample should be 150 μL.
4. Run sensorgrams of the six samples and one control (follow **steps 3–7**). The association phases of these sensorgrams will be used to evaluate the association rate constant $k_{on}$.

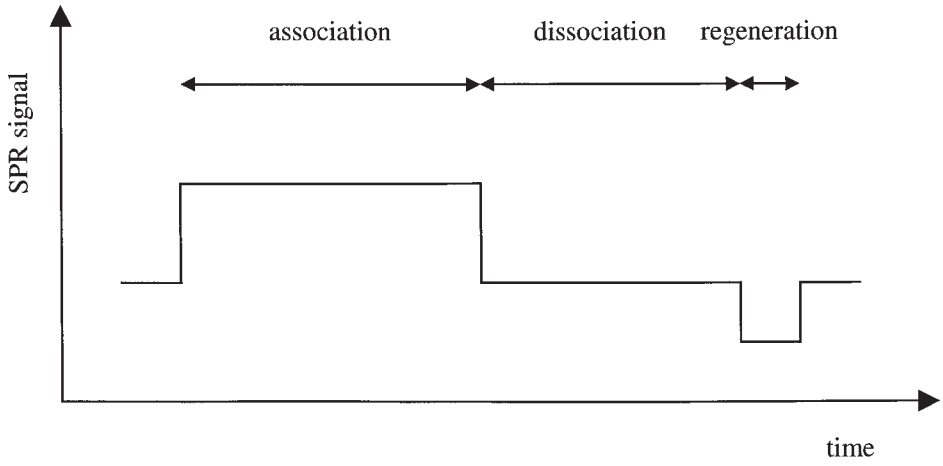


Fig. 3. A typical sensorgram when the analyte does not associate to the immobilized ligand. The rise in signal during the association phase is an effect of the bulk concentration change.

5. Make three samples with the highest concentration (50 n*M* in the example in **step 3**). Each sample should be 150 μL.
6. Run sensorgrams of the three samples, but instead of routinely recording 30 min of the dissociation phase; record it long enough for the signal to drop significantly, preferably all the way back to the baseline level. For a strong interaction, this may take several hours. If the instrument is stable (**Note 12**) and readily available, it may in some cases be meaningful to record at least one of the samples for 24 h or more if the dissociation rate is slow. The dissociation phases of these sensorgrams will be used to evaluate the dissociation rate constant $k_{off}$.
7. For evaluation, either start the evaluation software supplied with the instrument, or export the data into the fitting software of your choice.
8. The dissociation phase must be evaluated first. Delete the first few minutes of the dissociation phase because they are influenced by the bulk concentration change. Fit the rest of the dissociation phase to the following equation in the case of a 1:1 complex (*see* the manual for other stochiometry):

$$R(t) = C \exp(-k_{off}\, t) + R_0 + R_{drift}\, t$$

$k_{off}$ is the dissociation rate constant, $t$ is the time, $C$ is $R(t = 0)$, $R_0$ is the baseline value at $t = 0$, and $R_{drift}$ is the linear slope of the baseline. The drift term ($R_{drift}\, t$) might be excluded if the dissociation rate is fast and the recording time for the dissociation therefore short (<1 h) or if the instrument is very stable. If the values of $k_{off}$ from the three different experiments vary within a factor of 2, they can be considered constant.

9. Once the dissociation rate constant has been established, the association rate constant may be evaluated. Determine the exact time of injection $t_0$. Delete the first few minutes of the association phase because they are influenced by the bulk concentration change. Fit the rest of the association phase to the following equation in the 1:1 complex case (*see*the manual for other stochiometry):

$$R(t) = R_0 + R_{eq}(1 - \exp[-([A]k_{on} + k_{off})\cdot(t - t_0)])$$

$k_{on}$ is the association rate constant and $R_0$ is the SPR value at time 0. $R_0$ is generally not the same as the baseline value, because of the sensitivity of the SPR to changes in solvent composition. It is assumed that the analyte concentration [*A*] is held constant in the flow cell during the entire association phase of a certain SPR experiment because the flow of the mobile phase is fast compared to the association and dissociation reactions. The SPR value approaches $R_{eq}$ when $t$ approaches infinity. $R_{eq}$ is equal to

$$R_{eq} = R_{max} \cdot [A] \cdot k_{on} / [A] \cdot k_{on} + k_{off}$$

$R_{max}$ is the maximum SPR response, which would be observed if there was analyte bound to all available immobilized molecules (some molecules may be immobilized, but not available to binding, e.g., because of unsuitable geometry of the immobilization). $R_{max}$ would, in principle, be constant for all association experiments run on a particular immobilization surface. Alas, instabilities in the instrument and the deterioration of the immobilized molecule with time and number of experiments may cause variations in $R_{max}$. If the values of $k_{on}$ from the six different experiments vary within a factor of 2, they can be considered constant.

## 4. Notes

1. The most versatile matrix for immobilization of macromolecules is the carboxymethylated dextran matrix, which allows immobilization via native -$NH_2$, -SH, -CHO, and -COOH groups. Other immobilization matrices for use in the BIAcore instruments are:
   a. Streptavidin-coated matrix for binding of biotinylated ligand;
   b. Flat hydrophobic surface for lipid coating and binding of membrane binding macromolecules;
   c. NTA-coated matrix for nickel chelation and binding of hisitdine-tagged ligand. Other matrices are under development by the vendors.
2. When it comes to commercial SPR instruments, the BIAcore instruments are the most widespread (www.biacore.com). They all use a continuous flow technique, where the analyte is dissolved in the flow buffer and this is the technique described in this chapter. A newcomer to the field is the KI1 instrument by BioTul (www.biotul.com). This is a cuvet-based instrument; i.e., does not support the continuous flow technique. KI1 will be released after the preparation of this manuscript (August 1999). A technique similar to SPR, called resonant mirror

*(2)*, is used in the Iasys instruments by Affinity Sensors (www.affinity-sensors.com). Although the technique of achieving a refractive index dependent signal is different, there are many similarities, especially when it comes to the applications to which the instrument can be used. This is also a cuvet-based instrument.

3. BIAcore manufactures many of the chemicals needed for immobilization and binding experiments themselves. To reduce costs, try buying the same chemicals from your regular dealer of chemicals.
4. When adding the surfactant to the flow buffer, you may have problems pipetting because it is very viscous. Press the button of a fast-pipet and leave in the bottle for a couple of minutes.
5. The bottle top filters are costly, but the problems associated with clogging of the flow in the SPR instrument justify the cost. The important issue is the filter and its tight fit to the bottle, not the fact that it is sterile. Therefore, in many instances it is possible to use the filter again for the same kind of buffer.
6. Most BIAcore models are constructed to allow for 2 or more flow cells on each sensor chip. The number of flow cells depends on the particular model. The flow can be directed to go through any of the flow cells. For most coupling matrices, only one immobilization experiment can be performed in each flow cell. This is true for the carboxymethylated dextran matrix chip. Because the sensor chips are expensive (the BIAcore chips contain gold), it is important to keep track of what has been immobilized in each flow cell.
7. For immobilization to be successful, make sure the ionic strength of the protein stock is not too high, preferably below 20 m*M* cation concentration. If lyophilized protein is used, dissolve it in distilled water and make the stock strong enough not to influence the pH of the immobilization mixture.
8. For qualitative estimations of binding, immobilization may well be as high as possible because sensitivity is gained, but for quantitative estimations of the kinetics, it is important that the immobilized molecules do not interact with or disturb each other. Therefore, it is a good idea to try surfaces with different amounts of immobilized protein and compare the results. If there is no difference, one may trust the results obtained from the denser surface.
9. The sensitivity of SPR is highly dependent on the model of the instrument. Furthermore, it is important to remember that the signal always is proportional to the total weight of the molecules close to the surface. This means that experiments involving proteins with MW<5000 requires a very sensitive instrument and/or very high concentrations. If the two proteins involved in an interaction are very different in size, it is better to immobilize the smaller ligand to gain sensitivity.
10. If no association at all is found, first try higher analyte concentration. If that does not help, try immobilizing under different conditions to achieve higher surface density. It is also possible that the immobilized ligand is ruined by the regeneration procedure. Try other regeneration conditions to overcome this problem.
11. If the flow rate is too low, this may cause mass transport problems at the chip surface, which distorts the kinetics. However, for some instruments the injection

volume is maximized to 45 μL. If the flow is >10 μL/min, this means we can only monitor the association phase during 4.5 min or less, which is often too little. It is always safest to try different flow rates to see that this does not change the resulting rate constants.

12. Running dissociation experiments for extended times puts the stability of the instrument to a test. Slow baseline drifting is common and it is always safest to include a linear term in the fitting procedure of the dissociation rate constant. Spikes often occur in the sensorgram. They are caused by air bubbles in the flow and are minimized by proper degassing of the flow buffer. An occasional spike can be removed manually from the data before evaluation. More serious than drifting or spikes is clogging. The thin tubing of the instrument and the flow cells may be clogged and this causes the signal to flip out completely. First aid is to try the declogging routines described in the manual, but it is much better to prevent clogging altogether by routinely cleaning the instrument (follow the maintenance instructions in the manual).

## References

1. Kretschmann, E. and Raether, H. (1968) Radiative decay of non radiative surface plasmons excited by light. *Z. Naturforsch.* **23a,** 2135–2136.
2. Cush, R., Cronin, J. M., Stewart, W. J., Maule, C. H., Molloy, J., and Goddard, N. J. (1993) The resonant mirror: a novel optical biosensor for direct sensing of biomolecular interactions. Part I: Principle of operation and associated instrumentation. *Biosensors Bioelectronics* **8,** 347–353.

# 9

# Differential Scanning Calorimetry

**Maria M. Lopez and George I. Makhatadze**

## 1. Introduction

Protein folding/unfolding reaction, as any other chemical reaction, is accompanied by heat effects. The heat of unfolding measured at a constant pressure represents the enthalpy of the process. Direct measurements of the heat of unfolding are done using differential scanning calorimetry (DSC).

DSC measures the heat capacity of protein in aqueous solution as a function of temperature. The area under the heat-capacity profile represents the enthalpy of unfolding, the temperature of the maximum heat-capacity profile provides with the transition temperature, and the difference in the heat capacities of the native and unfolded states defines the temperature dependence of the enthalpy and entropy functions and, thus, the temperature dependence of protein stability. In addition to these parameters, DSC provides a direct estimate of the modes of protein unfolding. The sharpness of the heat-capacity profile gives another characteristic of the observed process, the effective enthalpy of transition, usually referred to as the van't Hoff enthalpy. The ratio of these two enthalpies provides information about the mode of the observed transition. A ratio equal to 1 indicates that the observed transition is two state, proceeding from the native to the unfolded state without a significant population of intermediates (*see* **Fig. 1**). Deviation from unity indicates that the transition is more complicated. The temperature dependence of the enthalpy of transition allows the estimate of the partition function of the system and makes it possible to deconvolute the DSC-profiles into individual transitions (*see* **Fig. 2**).

DSC operates in differential mode, which means that the heat capacity of the protein in aqueous solution is measured relative to the heat capacity of buffer. Ideally, when the heat capacities and volumes of both sample and reference cells are identical, single protein/buffer scan will suffice. However, in

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

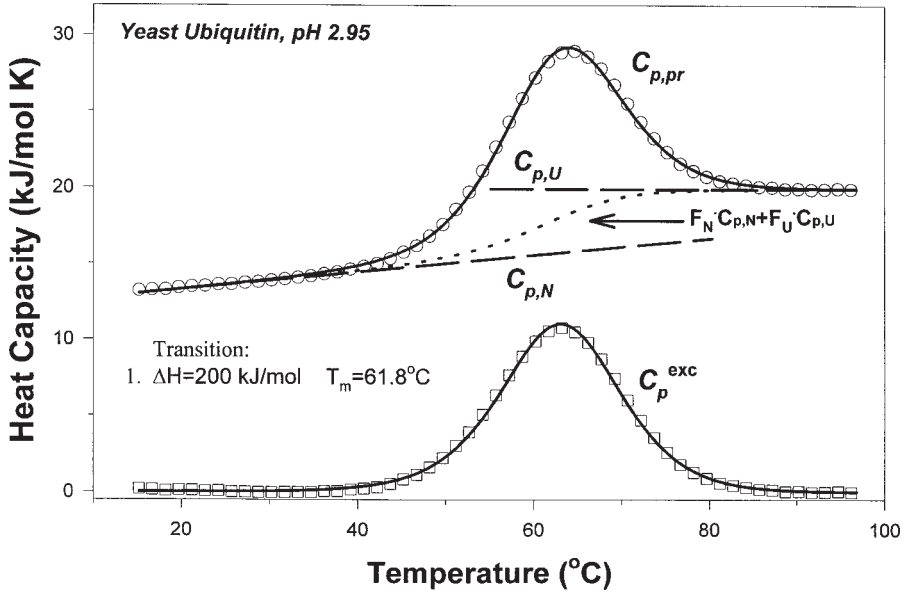


Fig. 1. The partial molar-heat capacity of ubiquitin at pH 2.95 obtained experimentally (O) and fitted according to a two-state model (*solid line*), the partial molar heat capacities of the native $C_{p,N}$ and unfolded states $C_{p,U}$ (*dashed lines*), the progress heat capacity $F_N \cdot C_{p,N} + F_U \cdot C_{p,U}$ (*dotted line*), the excess heat capacity $C_p^{exc}$ experimental (□) and fitted (*solid line*).

reality, sample and reference cells are slightly different and this difference has to be taken into consideration by recording buffer/buffer scan. Thus prior to starting the protein scan, it is very important to establish the stability of the baseline, i.e., its relative position and shape.

One of the most important user-defined parameters is the heating rate. Several considerations must be taken into account. First, the increase in sensitivity is linear with respect to the heating rate, i.e., sensitivity with a heating rate of 120°/h is twice higher than at 60°/h. One needs to keep in mind that the increase in sensitivity actually leads to the decrease in the signal-to-noise ratio. Second, if the expected transition is very sharp, e.g., occurs within a few degrees, a high heating rate will distort the shape of the heat absorption profile and lead to an error in the determination of all thermodynamic parameters for this transition and, in particular, the transition temperature. Third, the higher the heating rate, the less time the system has to relax to the equilibrium. For slow unfolding/refolding processes, it is preferred to use low heating rates. Usually small globular proteins exhibit fast folding/unfolding and the heating rates of 90–120°/h are acceptable. For larger proteins, it is

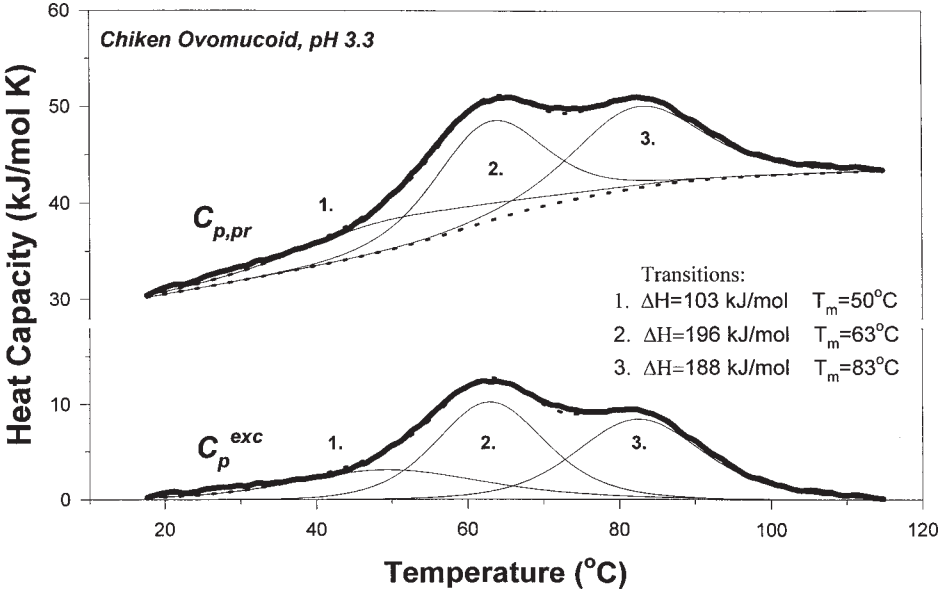


Fig. 2. The partial molar-heat capacity of chicken ovomucoid at pH 3.3 obtained experimentally (*thick solid line*) deconvoluted into three independent transitions (*thin solid lines*) with the thermodynamic parameters for individual transitions shown.

customary to use lower heating rates (30–60°/h). Fibrillar proteins such as collagen or myosin, which exhibit very narrow transitions, a heating rate of 10–20°/h is most suitable.

## 2. Materials

1. The DSC instrument designed to study biological systems must be extremely sensitive and require small amounts of the material, 0.1–1 mg/mL of protein solution. Currently there are two commercial DSC instruments, Nano-DSC from Calorimetric Science Corporation (Provo, UT) and VP-DSC from Microcal Inc. (Northhampton, MA). Both of these instruments are fully automated for control, data collection, handling, and analysis using PC computers. VP-DSC appears to have higher sensitivity and better baseline stability and is supplied with the superb ORIGIN graphics software.
2. Syringe with precut needle used to wash the cell and load the sample (provided by the manufacturer).
3. Spectrophotometer and quartz cells of different path length, for those cases in which the protein concentration is determined spectrophotometrically.
4. Dialysis bags with molecular weight cutoff depending on the molecular weights of the protein to be dialyzed.
5. Highly pure protein sample.

6. Buffers should be chosen for their low ionization enthalpy (and thus low-temperature dependence of p$K_a$) such as glycine (pH 2.0–3.5), sodium acetate (pH 3.5–5.0), sodium cacodylate (pH 5.5–7.0), sodium phosphate (pH 6.0–7.5).

## 3. Methods

### 3.1. Instrument Preparation

1. The instrument should be turned on at least 12 h prior to the experiment and "thermal history" established by running the baseline scans with the cells filled with the buffer.
2. Calibration of the instrument should be done periodically (once a year) using the procedure provided by the manufacturer (*see* **Note 1**).

### 3.2. Sample Preparation

1. Purified protein should be extensively dialyzed (with several changes of buffer every 6 h or more) against corresponding buffer.
2. Prior to the experiment insoluble particle and dust should be eliminated by centrifugation at 13,000*g*. Filtration of the protein solution is not recommended.
3. Measure protein concentration (*see* **Note 2**).

### 3.3. Data Collection

1. Thoroughly wash both cells with buffer from the last dialysis and fill them with the buffer. It is important to avoid any air bubbles trapped in the cell. For this, after filling up the syringe, pump out all air bubbles. Insert the needle into the calorimetric cell (needles are precut to a specific length so that the tip of the needle is barely above the bottom of the cell) and slowly lower the plunger until the solution appears in the overflow reservoir. At this point, start abruptly pumping solution in and out of the cell. This abrupt pumping will force trapped air bubbles out from the cell.
2. Fill both cells with the buffer and run buffer/buffer scan.
3. Refill the sample cell with the protein solution and run protein/buffer scan.
4. Rescan to check the reversibility of the unfolding (*see* **Note 3**).

### 3.4. Data Analysis

1. Subtract the protein/buffer scan from the buffer/buffer scan to obtain $\Delta C_p^{app}(T)$, the heat-capacity difference between sample and reference cells at temperature T.
2. Convert $\Delta C_p^{app}(T)$ into the partial heat capacity of the protein at temperature $T$, $C_{p,pr}^{\exp}(T)$ as:

$$C_{p,pr}^{\exp}(T) = C_{p,H_2O} / \overline{V}_{H_2O} \cdot \overline{V}pr - \Delta C_p^{app}(T) / m_{pr}$$

where $\overline{V}pr$ is the partial volume of the protein, $m_{pr}$ is the mass of the protein in the calorimetric cell, $\overline{V}_{H_2O}(T)$ is partial molar volume of aqueous buffer, and $C_{p,H_2O}$ is the heat capacity of aqueous buffer. The partial volume of the protein, $\overline{V}pr$ can be calculated from the amino acid composition of the protein using an

additivity scheme as described ***(1)***. The parameter $C_{p,H_2O} / \overline{V}_{H_2O}$ can be considered independent of temperature and equal to 4.2 J/(K·cm$^{-3}$).

3. Depending on the protein, the partial specific heat capacity of the native state $C_{p,N}$ at 25°C ranges from 1.25 to 1.80 J/K·g ***(2)***. The dependence of $C_{p,N}$ on temperature appears to be a linear function of temperature with a slope from 0.005 to 0.008 J/K$^{-2}$ g also depending on the protein ***(2)***. The partial specific heat capacity of the unfolded state $C_{p,U}$ is always higher than the heat capacity of the native state. At 25°C, $C_{p,U}$ values for different proteins range from 1.85 to 2.2 J/K·g, whereas at 100°C, $C_{p,U}$ values are higher, from 2.1 to 2.4 J/K·g ***(2)***. Partial heat capacity of the unfolded state has a nonlinear dependence on temperature (e.g., **ref. *3***). It increases gradually (with the slope comparable to that for the native state) and approaches a constant value at 60–75°C. The heat-capacity change upon protein unfolding, $\Delta C_p = C_{p,U} - C_{p,N}$, appears to be a temperature-dependent function. However, this dependence is weak in the temperature range 0–70°C, so in a first approximation, $\Delta C_p$ can be considered constant.
4. Analysis of the DSC profiles according to a certain model can be done using ORIGIN software from Microcal Inc. Alternatively, any nonlinear regression software (e.g., NONLIN, NLREG, SigmaPlot) can be used to write user defined scripts ***(4)***. An overview of the analysis of the complex non-two-state transitions is available ***(5)***. The following formalism is to be used for the simplest case when the unfolding is a monomolecular two-state process (*see* **Fig. 1**).

   The heat capacity functions for the native and unfolded states are represented by the linear functions of temperature, $T$ expressed in Kelvin, as ***(6)***:

$$C_{p,N}(T) = A_N \cdot (T - 273.15) + B_N \tag{1}$$

$$C_{p,U}(T) = A_U \cdot (T - 273.15) + B_U \tag{2}$$

The equilibrium constant of unfolding reaction, $K$, is related to the Gibbs energy change upon unfolding as:

$$K = \exp(-\Delta G / RT) \tag{3}$$

The Gibbs energy of unfolding, $\Delta G$, is defined as

$$\Delta G = T_t - T / T_t \cdot \Delta H_{\text{fit}}(T_t) + \Delta C_p \cdot (T - T_t) + T \cdot \Delta C_p \cdot \ln(T_t/T) \tag{4}$$

where $\Delta C_p$ is the heat-capacity change upon unfolding taken to be independent of temperature, $T_t$ is the transition temperature, and $\Delta H_{\text{fit}}(T_t)$ is the enthalpy of unfolding at $T_t$. The transition temperature is defined as the temperature at which the populations of the native $F_N$ and unfolded $F_U$ proteins are equal. The populations are defined by the equilibrium constant as:

$$F_N(T) = 1 / 1+K \quad \text{and} \quad F_U(T) = K / 1+K \tag{5}$$

The experimental partial molar heat-capacity function $C_{p,pr}(T)$ is fitted to the following expression:

$$C_{p,pr}(T) = F_N(T) \cdot C_{p,N}(T) + C_p^{exc}(T) + F_U(T) \cdot C_{p,U}(T) \tag{6}$$

The excess heat capacity defined $C_p^{exc}(T)$ as:

$$C_p^{exc}(T) = [\Delta H(T)^2 / R \cdot T^2] \cdot [K / (1+K)^2] \quad (7)$$

where the enthalpy function is defined as

$$\Delta H(T) = \Delta H_{fit}(T_t) + \Delta C_p \cdot (T - T_t) \quad (8)$$

There are seven fitted parameters: $T_t$, $\Delta H_{fit}$, $\Delta C_p$, $A_N$, $A_U$, $B_N$, and $B_U$.
In order to analyze the data according to these equations, the reversibility of unfolding reaction should be established experimentally by reheating the sample. If more than 80% of the original signal is recovered the reaction can be considered as reversible. For the analysis of the irreversible transitions, *see* **ref. *7***.

## 4. Notes

1. The DSC cell should be cleaned regularly. This can be accomplished in mist cases by filling the cells with 10% sodium dodecyl sulfate (SDS) and heating it up to 100°C, followed by a thorough rinse with distilled water. Alternatively, the cells can be washed with 200 proof ethanol followed by a wash with distilled water. Drying the cells is not recommended.
2. The protein concentration is a very important parameter because it is required for the quantitative analysis according to **Eqs. 1–8**. The extinction coefficient can be calculated from the number of aromatic residues and disulfide bonds in a protein using an empirical equation (**ref. *8***):

$$\varepsilon_{280\,nm}^{0.1\%,1cm} = (5690 \cdot N_{Trp} + 1280 \cdot N_{Tyr} + 120 \cdot N_{SS}) / Mw \quad (9)$$

   where *Mw* is the molecular mass of the protein in daltons. A simple experimental procedure for estimating the extinction coefficient is described ***(9)***.
3. The reversibility of unfolding strongly depends on the upper temperature limit during the first scan.

## References

1. Makhatadze, G. I., Medvedkin, V. N., and Privalov, P. L. (1990) Partial molar volumes of polypeptides and their constituent groups in aqueous solution over a broad temperature range. *Biopolymers* **30,** 1001–1010.
2. Makhatadze, G. I. (1998) Heat capacities of amino acids, peptides and proteins. *Biophys. Chem.* **71,** 133–156.
3. Makhatadze, G. I. and Privalov, P. L. (1995) Energetics of protein structure. *Adv. Prot. Chem.* **47,** 307–425.
4. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) Thermal vs guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **38,** 8138–8149.
5. Biltonen, R. L. and Freire, E. (1978) Thermodynamic characterization of conformational states of biological macromolecules using differential scanning calorimetry. *CRC Crit. Rev. Biochem.* **5,** 85–124.

6. Makhatadze, G. I. (1998) *Measuring Protein Thermostability by Differential Scanning Calorimetry*, vol. 2, Wiley, New York.
7. Sanchez-Ruiz, J. M. (1992) Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophys. J.* **61,** 921–935.
8. Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319–326.
9. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423.

# 10

# Isothermal Titration Calorimetry

**Maria M. Lopez and George I. Makhatadze**

## 1. Introduction

Thermodynamic characterization of the system provides us with important information about the stability, strength, specificity, and stoichiometry of interacting systems. The method of choice for the direct measurements of energetics of protein-ligand interactions is the isothermal titration calorimetry (ITC) technique. The use of ITC to measure the binding of a macromolecule (in general, a protein) to its ligand (ion, peptide, another protein, DNA, RNA, and so on) relies on the fact that such an interaction is accompanied by a heat effect. The heat absorbed (endothermic) or released (exothermic) upon the interaction $Q$ is then used to obtain information about the binding constant, $K_a$, and the enthalpy of binding $\Delta H$. The strength of ITC is that under proper experimental conditions, from one single titration both the binding constant and the enthalpy of binding can be obtained. Moreover, the temperature dependence of the enthalpy of binding allows one to calculate another important thermodynamic parameter: the heat capacity change of binding $\Delta C_p$. The $\Delta C_p$ is calculated from the slope of $\Delta H$ vs temperature, and it can be positive (hydrophobic interactions are disrupted upon binding) or negative (hydrophobic interactions are formed) ***(1)***.

The shape of the binding curve depends on the unitless parameter $c$. This parameter is proportional to the equilibrium constant $K_a$ the total concentration of protein in the cell $[P_t]$ , and the stoichiometry of the interaction $n$, as:

$$c = K_a \cdot [P_t] \cdot n \tag{1}$$

High $c$ values indicate very tight binding and low $c$ values indicate low affinity of the interactions. High-affinity binding constants have to be measured at low protein concentration. However, one needs to keep in mind that the lower the

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

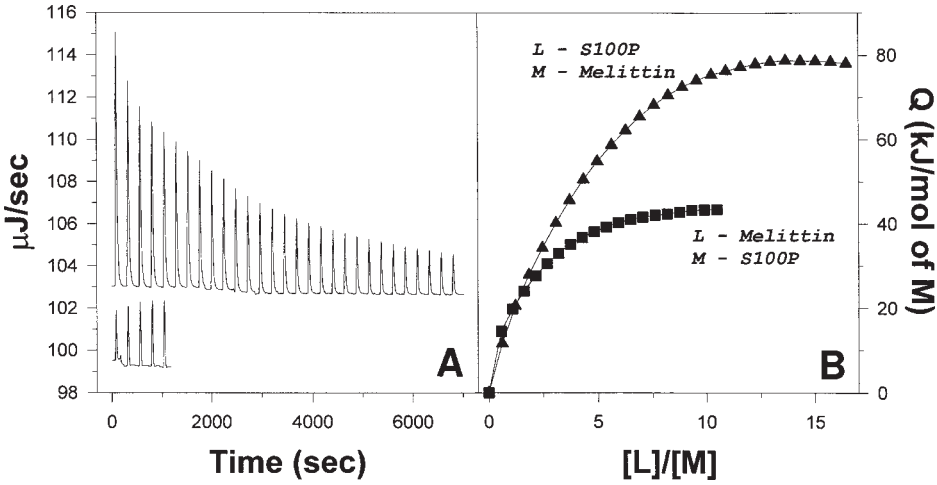


Fig. 1. Isothermal titration calorimetric study of the interactions of $Ca^{2+}$-binding protein S100P with the 29 amino acid residue peptide melittin (Guzman-Casado, M. and Makhatadze, G. I., personal communication). **(A)** Experimental profile of 29 injections of 10 μL each of melittin (1.9 m*M*) into the cell containing 22 μ*M* S100P obtained at 32°C using VP-ITC (upper profile). The heat of dilution of melittin was measured by injecting 10 μL of melittin (1.9 m*M*) into the cell containing just the 50 m*M* Tris-HCl, 2 m*M* EDTA, pH 7.5 buffer (lower profile). Note that the peaks corresponding to the melittin injection into the buffer are similar to the peaks observed at the last several injections of melittin into S100P. **(B)** The integral heat of interactions (normalized per mole of melittin) between S100P and melittin obtained from the experiment shown on **(A)**, i.e., injection of S100P into melittin are shown as solid triangles (▲). In the parallel experiment, mellitin was injected into the solution of S100P (actual data not shown) and the integral heat effect (normalized per mole of S100P) is shown as solid squares (■). If the stoichiometry of interactions would have been 1:1 both experiments would overlap. Otherwise the ratio of the heat effects at the saturations will give the stochiometry of the interactions which in the shown example is two molecules of melittin per molecule of S100P, with the equilibrium constant $K_d \approx$ 80 μ*M*.

protein concentration in the in the cell, the weaker the signal. This means that there is an upper limit in the values of the affinity constants that can be measured. For example, the sensitivity of VP-ITC (Microcal Inc.) is 0.1 μcal and accurate measurements require 3–5 μcal per injection into 1.3-mL cell. Thus, binding constants in the order of $10^7$ $M^{-1}$ can be easily obtained for reactions with $\Delta H = -10$ Kcal/mol (or even higher if $\Delta H$ is larger). For the binding reactions with very high affinity ($\geq 10^9$ $M^{-1}$) the equilibrium constant cannot be measured and only the enthalpy of binding, $\Delta H$, can be estimated (*see* **Fig. 1**).

Most of the binding reactions with one binding site have $\Delta H$ in the range from 4 kcal/mol to –25 kcal/mol. In the case of very low-affinity binding constants, saturation may never be achieved and thus the equilibrium constant and enthalpy cannot be estimated directly from the fit of the data. The lower limit for $K_a$ is approx $10^3\ M^{-1}$.

## 2. Materials

1. Isothermal titration calorimeter, such as VP-ITC (Microcal Inc., Northhampton, MA).
2. Syringe with precut needle used to wash the cell and load the sample (provided by the manufacturer).
3. Stirring syringe with precut needle used for the ligand (provided by the manufacturer).
4. Spectrophotometer and quartz cells of different path length, for those cases in which the protein and ligand concentration are determined spectrophotometrically.
5. Dialysis bags with molecular weight cutoff depending on the molecular weights of the protein and ligand to be dialyzed.
6. Highly pure protein and ligand are required as small contaminations will carry errors in their concentrations and will make all binding parameters meaningless.
7. Buffers should be chosen based on the solubility and stability of the protein and ligand and on the chemical compatibility with the cell material. Additives such as 1,4,-dithiotreitol (DTT) are not recommended to use in VP-ITC and should be substituted with tris-(carboxyethyl)phosphine (TCEP).

## 3. Method

### 3.1. Instrument Preparation

1. The day before starting the experiments, set the desired experimental temperature of the calorimeter. This is particularly important if the experiment is to be performed at low temperature, because the ITC needs time to cool down. When experiments at different temperatures are planned start with the titration at the lowest $T$ because it is much faster to heat up the ITC cells than to cool them down (*see* **Note 1**).
2. Calibration of ITC instruments should be done periodically (at least once a year) using the procedure provided by the manufacturer.

### 3.2. Sample Preparation

1. Proper dialysis of the protein and ligand is a key point in order to get reliable results. Dialyze protein and ligand at the same time in the same buffer. Three changes, one every 6 h, in general, is enough. If the ligand is too small to dialyze, dissolve it directly in the last dialysis buffer. When preparing the solutions, take into account that it requires 50–80% larger volume of protein solution to fill up the cell. For example, the volume of the VP-ITC cell is approx 1.5 mL, but approx

2.5 mL are required to it fill up. Similarly, to fill a 250-μL stirring syringe for VP-ITC, it is necessary to prepare approx 400 μL of ligand solution.

2. The protein concentration in the cell can be very low, and thus it may be advantageous to prepare a stock solution of protein. It is crucial that the dilution of the stock protein solution to the desired concentration is done with the buffer used for the last dialysis. This will save time when experiments at different concentrations are required or when different experimental conditions are tested (*see* **Note 2**).
3. Remove the protein and ligand solutions from the dialysis bag and centrifuge them for at least 20 min at 14,000 rpm and 4°C to remove insoluble particles and dust. Never filter solutions because there might be nonspecific absorption of protein to the membrane!
4. Measure concentrations of the protein and ligand. Knowledge of exact concentrations of protein and ligand are very important. Rapid and accurate method for measuring concentration is optical absorbance in the UV-range. The extinction coefficient can be calculated from the number of aromatic residues and disulfide bonds in a protein using an empirical **Eq. 2**:

$$\varepsilon_{280\text{nm}}^{0.1\%,1\text{cm}} = (5690 \cdot N_{\text{Trp}} + 1280 \cdot N_{\text{Tyr}} + 120 \cdot N_{\text{SS}}) / Mw \quad (2)$$

where $Mw$ is the molecular mass of the protein in daltons, and $N_{\text{Trp}}$, $N_{\text{Tyr}}$, and $N_{SS}$ are number of tryptophan, tyrosine, and disulfide bonds, respectively. A simple experimental procedure for estimating the extinction coefficient is described ***(3)***. In many cases, the correction for light scattering is required and can be taken into account according to ***(4)***.
5. Wash the cell with the buffer from the last dialysis. Introduce the needle very carefully in the cell until it touches the bottom. At this time, lift it up about 1 mm from the bottom and remove the liquid in the cell. The cells should never be dried! Inject the buffer solution until you see it overflowing at the top (in the reservoir area) and then remove the buffer from the cell. Repeat this procedure 20–25 times (approx 50 mL of buffer). After last washing, remove all buffer left in the cell. Remove all remaining buffer from the washing syringe by gently shaking it, but do not dry.
6. Fill the cell with the protein solution. It is important to avoid any air bubbles trapped in the cell. For this, after filling up the syringe, pump out all air bubbles. Insert the needle into the calorimetric cell (so that the tip of the needle is barely above the bottom of the cell) and slowly lower the plunger until the solution appears in the overflow reservoir. At this point, start abruptly pumping solution in and out of the cell. This abrupt pumping will force trapped air bubbles out from the cell.
7. The ITC reference cell is generally loaded with 0.1% sodium azide solution in water and needs to be changed only periodically (once a month).
8. Fill up the ITC stirring syringe (the VP-ITC comes with special narrow tubes necessary to load the stirring syringe). Once the syringe is loaded, make sure there are no bubbles, wipe the needle of the syringe with a Kimwipe towel. Avoid touching the wholes in the paddle.

9. Place the stirring syringe very carefully in the ITC sample cell keeping it in vertical position at all times.
10. Start running the program for the titration. At least 15–20 injections of 3 to 10 μL each are recommended.
11. When the titration is done, remove the stirring syringe very carefully to avoid bending the needle. Empty the ITC sample cell and wash it with at least 50 mL of distilled water. Remove any ligand solution from the stirring syringe and wash it with at least 20 mL of distilled water and air-dry it (*see* **Note 3**).

### 3.3. Data Collection

The data is collected according to the program provided by the manufacturer. Before starting the experimental injection the instrument needs to be equilibrated. Equilibrium is reached when the signal coming from the instrument remains stable. At this time injection of the ligand into the sample cell can be started. The titration will continue until no more changes in the heat effects are observed, which will indicate that the protein is completely saturated with ligand.

1. Perform titration of protein with the ligand.
2. Perform titration of buffer only with the same ligand. This will allow one to take into account possible heat effect of dilution of the ligand.
3. Integrate heat effect of each injection for protein/ligand titration and correct by the heat effect of buffer/ligand titration. **Figure 1** shows a typical titration curve.

### 3.4. Data Analysis

1. *Single binding site*. This is the simplest model possible in which only one molecule of ligand $L$ binds to one molecule of protein $P$:

$$P + L \leftrightarrow PL$$

The heat absorbed or released upon the interaction $Q$ is proportional to the ligand bound $[L]_b$

$$Q = \sum_i Q_i = \Delta H \cdot V \cdot [L]_b \tag{3}$$

where $Q_i$ is the heat absorbed during an $i$th injection, $\Delta H$ is the enthalpy of binding and $V$ is the volume of the ITC cell. The concentration of ligand bound is given by

$$[L]_b = [P]_t \cdot (K_a \cdot [L]_f / 1 + K_a \cdot [L]_f) \tag{4}$$

where $[P]_t$ is the total concentration of the protein and $[L]_f$ is the free-ligand concentration ($[L]_f = [L]_t - [L]_b$). Considering these expressions, a final equation of $Q$ as a function of total protein and ligand concentrations can be obtained. The binding parameters $\Delta H$ and $K_a$ are obtained from the fit of $Q$ as a function of $[L]_t$. Another way of analyzing the binding reaction is considering the sigmoidal de-

pendence of Q on $[L]_t/[P]_t$ ***(5)***. This approach is used by the VP-ITC ORIGIN software and presents certain advantages when complicated binding models need to be used.

2. Identical and noninteracting binding sites. In this case more than one molecule of ligand binds to one molecule of protein ($n$ >1). The **Eq. 3** now becomes:

$$Q = {}_i\sum Q_i = n \cdot \Delta H \cdot V \cdot K_a \cdot [L]_b \quad (5)$$

and can be calculated as described by **Eq. 3**.

3. Nonidentical and noninteracting binding sites. This is the most general case in which there are "$i$" different sets of binding sites and each set has its own number of binding sites $n_i$, enthalpy $\Delta H_i$, and equilibrium constant $K^i_a$. The heat involved in the binding process can be written as:

$$Q = \sum_i Q_i = V \cdot \sum_i n_i \cdot \Delta H_i \cdot \mathrm{V} \cdot K^i_a \cdot [L]^i_b \quad (6)$$

and nonlinear regression allows estimated of the binding parameters to the each type of the binding site.

## 4. Notes

1. To avoid long equilibration times, keep the washing buffer and protein solution at a lower temperature than the temperature at which you want to perform the titration.
2. Keep always the buffer from the last change of the dialysis under the same conditions as your protein and ligand. Even if the protein and ligand are stable at 4°C, try to use them as soon as possible to ensure that the properties of the samples and buffer remain the same. If samples are not used within several days, consider re-dialysis with one change of buffer.
3. Special care has to be taken with the needle from the stirring syringe. Any small bending of the needle will disturb the baseline of the instrument and will make the syringe unusable.

## References

1. Makhatadze, G. I. (1998) Heat capacities of amino acids, peptides and proteins. *Biophys. Chem.* **71,** 133–156.
2. Gill, S. C. and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319–326.
3. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423.
4. Winder, A. F. and Gent, W. L. (1971) Correction of light-scattering errors in spectrophotometric protein determinations. *Biopolymers* **10,** 1243–51.
5. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179,** 131–137.

# 11

# Multiangle Laser Light Scattering and Sedimentation Equilibrium

**Leslie D. Hicks, Jean-René Alattia, Mitsuhiko Ikura, and Cyril M. Kay**

## 1. Introduction

Multiangle laser light scattering (MALLS) and sedimentation equilibrium are two powerful techniques used to characterize the association properties of proteins and their interactions in solution under physiological conditions. Both techniques have undergone a resurgence as a result of the advent of recombinant technologies which has enabled the generation of reasonable quantities of biologically significant proteins that exist in vivo in small amounts so that they can now be characterized physicochemically. As well, new technical developments with both techniques have made them much more sensitive and user friendly. In the case of static light scattering, this includes the use of lasers and modern detectors on-line with size exclusion chromatography so that one can establish absolute molecular weights of individual protein fractions eluting from the column. With sedimentation equilibrium, the Optima XL-I centrifuge (developed by Beckman, Palo Alto, CA) is equipped with both a new photoelectric scanning absorption optical system enabling exact measurement of concentration profiles at wavelengths of 190–800 nm and an interference optical system allowing the measurement of much higher concentration gradients. In both cases, powerful computer programs have been developed for data evaluation. Examples of the use of both techniques to study the association properties of cadherin in the presence and absence of calcium are described later.

## 2. MALLS

### *2.1. Instrument Set-Up and Preparation*

1. The instrumental setup consists sequentially of a solvent reservoir; Shimadzu model DGU-4A vacuum solvent degasser; Waters model 510 HPLC pump

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

equipped with a pulse dampener, in-line 0.2-μm and 0.1-μm membrane filters, and a Rheodyne Model 7125 sample injector complete with a 1.0-mL sample loop; Pharmacia Superose 12 FPLC Gel Filtration column; Wyatt Technology Corporation Dawn F Multi-Angle Laser Light Scattering Photometer equipped with a 10-mW 488-nm argon-ion laser and temperature-controlled flow cell and read head; Wyatt Technology Corporation Optilab Model 903 Differential Refractometer equipped with temperature-controlled 1-mm path length flow cell; and finally, a Lauda model K-2/R circulating water bath with insulated hoses connected to both the Dawn F and Optilab 903.

2. Periodically, the 90° detector in the Dawn F is calibrated with HPLC grade Toluene, which has a known scattering intensity, or Rayleigh Ratio, $R_\Theta$, following the procedures outlined in the Dawn F instrument manual ***(1)***. The Optilab 903 is calibrated with a solution prepared from ultrapure, anhydrous, NaCl following the procedures outlined in the Optilab 903 instrument manual ***(2)***.
3. The desired run temperature (usually 20°C.) is set for both the Dawn F and Optilab 903 using the temperature control setting on the Dawn F and on the Lauda water bath at least 24 h prior to calibrating or running samples to allow the Optilab signal (DRI) to stabilize, because it is very sensitive to temperature changes. As well, the Optilab lamp is powered up for a minimum of 24 h before data collection to reduce signal drift over time.
4. Prior to pumping through the instrument setup, the buffer solution is filtered through 0.1-μm membrane filters. A minimum volume of 100–200 mL is then pumped through the set-up at a maximum of 1.0 mL/min to allow the column to equilibrate and the Optilab signal to stabilize.

### 2.2. Sample Preparation

1. The protein solution (typically 400–500 μL of 1–5 mg/mL solution) is dialyzed against the buffer solution for 24 h and then filtered through a 0.22-μm syringe filter prior to injecting into the instrument set-up. After filtering, the sample solution is spun for 3–5 min in a microfuge to remove minute air bubbles introduced in the filtering process.

### 2.3. Data Collection

1. Prior to running a sample, the intensity of the scattering signal at different angles is normalized to the intensity of the signal at 90° in the Dawn F using a solution of BSA (Sigma 98% monomer), prepared in the same manner as the sample solution. With the HPLC pump operating at a flow rate of 0.8 mL/min, which provides reasonable separation of peaks and relatively quiet MALLS signals in reasonably short run times, a 40-μL aliquot of BSA solution (approx 5 mg/mL) is injected into the system using the Rheodyene injector. Data collection is performed using Wyatt Technology's Astra software.
2. Once the MALLS and DRI signals have stabilized after elution of the BSA peaks, an aliquot of the sample solution is injected (with the Superose 12 column and a sensitivity setting of 100, an injected mass of approx 0.25 mg will produce a full-

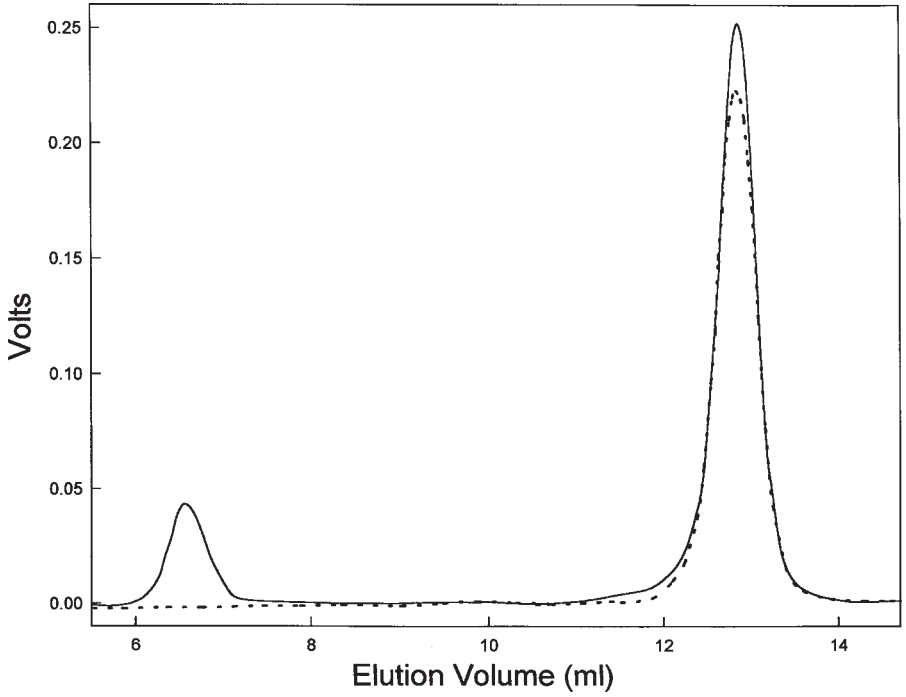


Fig. 1. Overlay of the 90° laser light scattering (———) and the DRI (- - - - -) chromatograms for E-cadherin in the absence of calcium. A 30-µL aliquot of a 7 mg/mL solution was injected at a flow rate of 0.8 mL/min. E-cadherin eluted as a single, narrow, symmetrical peak with a calculated weight average molecular weight of 24,370. The peak that appears at the void volume on the 90° chromatogram and is absent on the DRI chromatogram results from a pressure effect from the sample injection (*see* **Subheading 2.5., item 3**).

scale signal in the Optilab 903). Data is collected until the peaks of interest have eluted and the MALLS and DRI signals have returned to baseline levels.

### 2.4. Data Analysis

1. Data is analyzed using Wyatt Technology's Astra and Easi software following procedures outlined in the Dawn F instrument manual ***(1)*** and by Wyatt ***(3)***. The peak in the E-cadherin chromatogram (*see* **Fig. 1**) consists of many data slices collected in both the Dawn F and Optilab 903. Using either a measured or estimated value for the specific refractive increment, dn/dc (i.e., 0.185 for BSA), the concentration of each MALLS data slice is determined from the intensity of the corresponding Optilab data slice.
2. Combining this concentration value with the measured scattering intensities at the various angles in the Dawn F, a Debye plot ($R_\Theta/K^*c$ vs $\sin^2\Theta/2$), where $K^*$ is

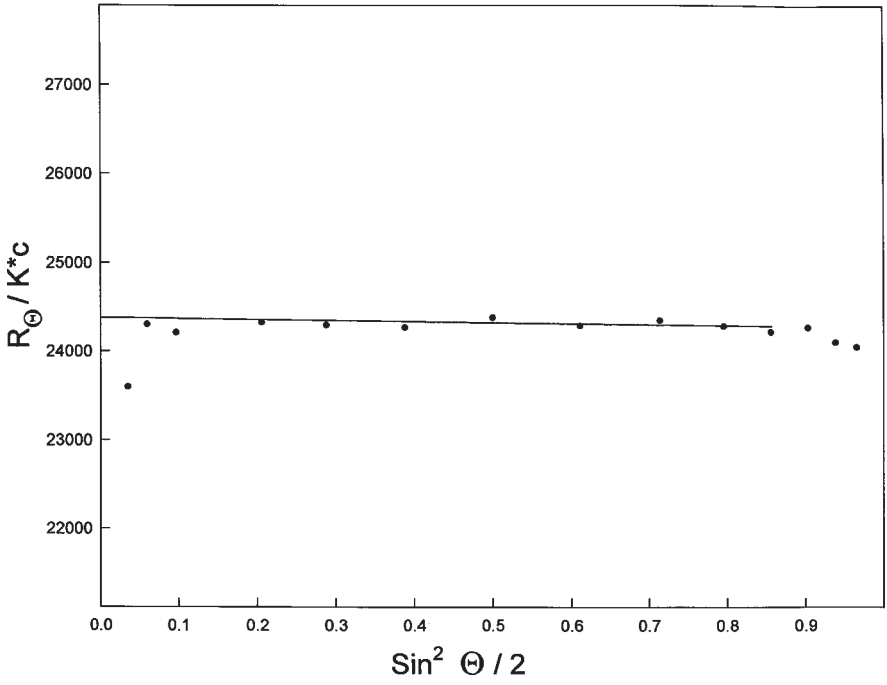


Fig. 2. Debye plot for one of the data slices from the E-cadherin peak illustrating the extrapolation of the scattering intensities to 0 angle. The calculated molecular weight for this particular data slice was 24,367.

a physical constant, and θ is the measuring angle, is produced for each data slice in the E-cadherin peak (*see* **Fig. 2**). In the Debye plot, the intercept of the extrapolation of scattering intensities to 0 angle yields the molecular weight directly. Once the molecular weight and concentration for each data slice has been determined, the number, weight, and *z* average molecular weights can be determined using the following equations:

Number Average: $M_n = \sum c_i / \sum (c_i/M_i)$
Weight Average: $M_w = \sum (c_i M_i) / \sum c_i$
*z* Average: $M_z = \sum (c_i M_i^2) / \sum (c_i M_i)$

Good agreement between the three calculated values indicates a fairly narrow, monodisperse molecular weight distribution across the peak of interest, whereas poor agreement indicates a more polydisperse distribution.

## 2.5. Notes

1. The flow cells in the Dawn F and Optilab 903 require utmost cleanliness for reliable data collection. The Dawn F flow cell can be removed periodically and disassembled for careful, manual cleaning, whereas the Optilab 903 flow cell

should not be disassembled, but can be flushed with cleaning solutions if necessary. However, the need for cleaning can be greatly reduced if the instrument setup is thoroughly flushed after sample runs with buffer solution and then MilliQ $H_2O$/0.01% $NaN_3$. Once the buffer solution is completely flushed from the setup, the HPLC pump is left running continuously at a low flow rate, circulating the MilliQ $H_2O$/0.01% $NaN_3$ through the column and flow cells and back to the solvent reservoir. This also extends the life of the various seals and check valves on the HPLC pump.

2. Careful filtering of both the buffer solution and sample solution and the use of an in-line vacuum solvent degasser cannot be emphasized enough, because the MALLS detectors are extremely sensitive to minute amounts of dust or micro air bubbles, which will increase the noise levels dramatically.
3. The MALLS detectors are very sensitive to pressure changes and a small peak invariably appears at the void volume on the MALLS chromatograms due to a pressure surge from the sample injection. This effect can be reduced by installing a bypass loop (3 feet of 0.01 in id PEEK tubing) using T-connections on either side of the Rheodyne injector.

## 3. Methods: Analytical Ultracentrifugation

### 3.1. Instrument Setup and Preparation

1. The instrumental setup consists of a Beckman XL-I Analytical Ultracentrifuge, containing both Interference and Absorbance Optical Systems, an IB*M* Pentium II computer for centrifuge control and data analysis, an AN 50 Ti 8-hole rotor, CFE six-sector sample cells, containing either quartz or sapphire windows, and a Perkin Elmer Lambda 5 dual beam spectrophotometer. Prior to performing sedimentation equilibrium runs on the samples at the desired run speeds, the radial calibration of the optical system being used is performed at low speed, typically 3000 rpm, following the procedures outlined in the XLI Instruction Manual ***(4)***.
2. The AN 50 Ti rotor is brought close to the desired run temperature, usually 20°C, prior to loading in the centrifuge to lessen equilibration time to the set temperature before the run start.

### 3.2. Sample Preparation

1. The protein solution, typically 400–500 μL of 1–5 mg/mL solution for use with the Interference Optical System, and a generally much lower concentration for use with the Absorbance Optical System, is dialyzed against the buffer solution for a minimum of 24 h and then both the dialyzate and protein solution are filtered through 0.22-μm syringe filters.
2. If there are aromatics present in the protein, the absorbance of the protein solution is measured against the dialyzate and an approximate concentration determined using a theoretical extinction coefficient calculated from the amino acid composition using the Sednterp computer program ***(5)***.
3. Dilutions are made of the stock solution to provide three loading concentrations, typically covering an approximate fivefold difference, such as 5.0, 3.0,

and 1.0 mg/mL for the Interference Optical System. Depending on the extinction coefficient of the protein and the wavelength being monitored in the XLI (optimum range of 230–600 nm), dilutions are made to provide approximately a fivefold difference in absorbance, ranging from 0.5–0.1 AU.

4. The six-sector centerpiece is placed in the outer ring of the sample cell. A window is placed in the lower window holder (sapphire for interference optics, quartz for absorbance optics), and the window holder is placed in the outer ring, followed by a cell-housing gasket and screw ring. The cell is placed in a torque-wrench assembly and the ring is tightened to 60-inch pounds. The cell is then removed from the wrench assembly and placed upright on the bench.
5. 110-μL of each loading concentration are pipeted into the right three sectors of the sample cell, with the highest concentration toward the top of the cell and the lowest concentration toward the bottom of the cell, so that once installed in the rotor, the lower concentrations will be at a larger radial distance from the rotor center and thus exposed to a higher centrifugal field. This ideally will provide optimum concentration gradients in each sector at each speed setting during the sedimentation equilibrium run.
6. 115 μL of dialyzate are loaded into the left three sectors for interference runs (125 μL for absorbance runs), the upper window and window holder, cell housing gasket, and screw ring are installed and then both screw rings are tightened to 130-inch pounds in the torque-wrench assembly. The top screw ring is tightened first to prevent sample leakage.
7. The sample cell(s) and appropriately weighted counterbalance are then placed in the rotor, and the rotor placed in the XLI, followed by the Monochromator/Laser Light Source assembly.
8. The vacuum system is then initiated and the rotor accelerated to 3000 rpm, at which time the radial calibration is performed as aforementioned while waiting for the rotor to equilibrate to the set temperature.

### 3.3. Data Collection

1. Once the radial calibration is completed and the rotor temperature has stabilized, the rotor is accelerated to the first run speed (typically two or three different speed settings are used for a sedimentation equilibrium run, ideally producing an approximate fourfold change in concentration from the meniscus to the cell bottom), and the rotor is left at this speed until sedimentation equilibrium has been attained.
2. Achievement of equilibrium is determined by taking successive scans at 2-h time intervals and subtracting one scan from another using the subtract utility in the Beckman analysis software. When there is no significant difference between two successive scans, equilibrium has been achieved and the last scan is saved for data analysis.
3. The rotor is then accelerated to the next speed and the process repeated until all the desired speed settings and equilibrium scans have been obtained.

### 3.4. Data Analysis

The sedimentation equilibrium data is evaluated using a nonlinear least-squares curve-fitting algorithm ***(6)*** contained in the NonLin analysis program ***(7)***. This program allows the analysis of both single and multiple data files. The self-association model shown in the following equation allows analysis of either a single ideal species or up to four associating species, depending on which parameters are permitted to vary during the fitting routine:

$$c_{\text{total}} = \delta c + \sum_{i=1}^{n} C_i\,(r) = \delta c + \sum_{i=1}^{n} K_{l,i} * C_1\,(r)^{q(i)}$$

where δc is the concentration offset of the first data point (this is generally 0 for absorption data and some arbitrary value for interference data); $q(i)$ is the degree of association for the $i$th associated species; $C_i(r)$ is the concentration of the $i$th species at radius $r$; $K_{1,i}$ is the equilibrium constant for the association of monomer to the $q(i)$- mer; $C_1(r)$ is the concentration of the monomer at radius $r$; $n$ is the total number of species present in the model being used for fitting the data, and can be between 1 (no association) and 5 for the current version of Nonlin.

The variable $C_1(r)$ may be expanded to

$$C_1\,(r) = C_{1,0} * \exp\left[\sigma * (\xi - \xi_0) - 2 * B * \sum_{i=1}^{n} C_i\,(r)\right]$$

where $C_{1,0}$ is the monomer concentration at the first point of the data set (the natural logarithm of this is used by the program as LnA); ξ, $\xi_0$ are the value of $r^2/2$ and $r^2/2$ at the first point of the data set (used as a reference position); $B$ is the colligative second virial coefficient with the assumption of $g_i = (g_1)^{q(i)}$; and σ is the reduced molecular weight as defined by

$$\sigma = M_1 * (1 - \bar{v} * \rho) * \omega 2 / RT$$

where $M_1$ is the monomer molecular weight; $\bar{v}$ is the partial specific volume; ρ is the solvent density; and ω is the rotor speed in radians per second.

**Figure 3** illustrates typical concentration vs $r^2/2$ plots and residuals vs $r^2/2$ plots for E-cadherin produced by the Nonlin program. The residual plots are useful for determining how well the data fits the selected model. A good fit should result in randomly scattered residuals with no apparent systematic patterns. As well, the square root of the variance should be less than $2.0 \times 10^{-2}$.

If a sample fits a multiple species model, the raw $K_a$ in absorbance or fringes, depending on the optical system being used, must be converted to a molar $K_a$ using the following equations:

$$K_{\text{conc}} = K_{\text{abs}} \times (Mx\ \wp)^{n-1} / n$$

where $K_{\text{conc}}$ is the association constant in molar concentration terms depending on the stoichiometry, $K_{\text{abs}}$ is the absorbance association constant from the fit,

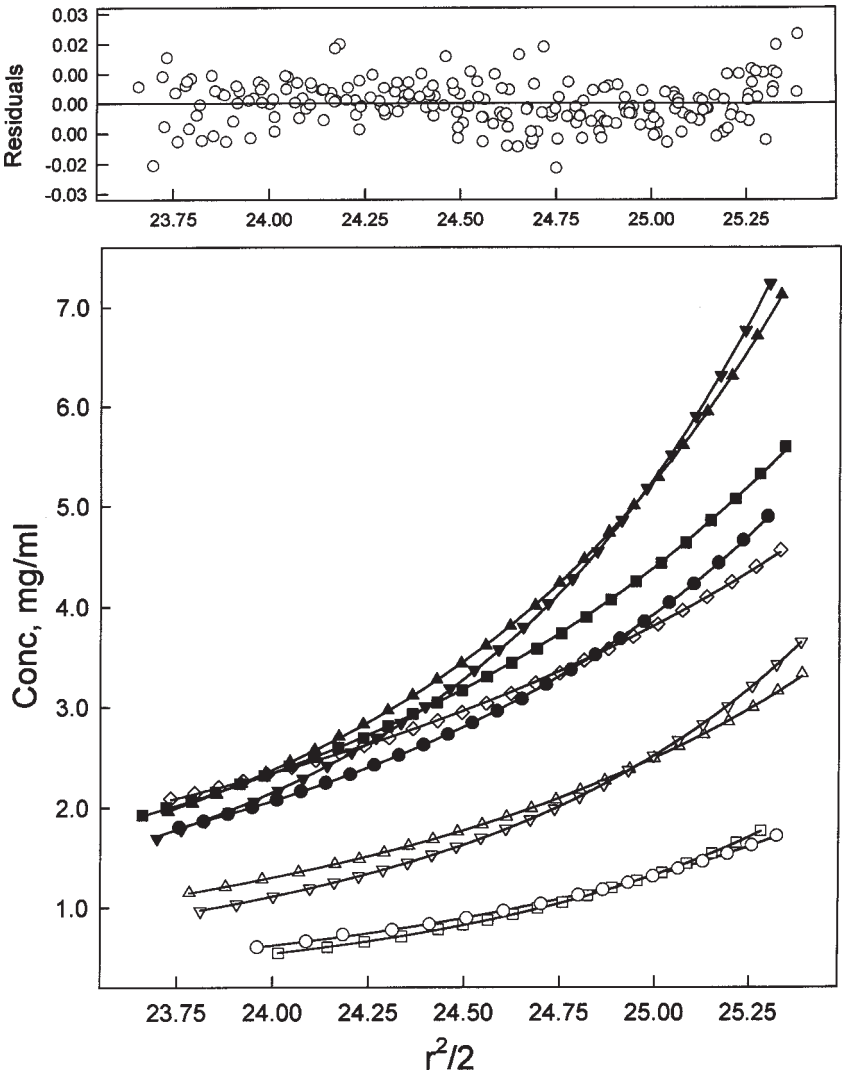


Fig. 3. Sedimentation equilibrium analysis of E-cadherin carried out in the presence of 10 m*M* $CaCl_2$. A total of nine runs were performed at different E-cadherin loading concentrations and operating speeds. The lower graph shows the local E-cadherin concentrations as a function of the radial distance: ○ 1.05 mg/mL (14,000 rpm), □ 1.05 mg/mL (16,000 rpm), △ 2.11 mg/mL (13,000 rpm), ▽ 2.11 mg/mL (15,000 rpm), ◊ 3.24 mg/mL (11,000 rpm), ● 3.24 mg/mL (13,000 rpm), ■ 3.62 mg/mL (12,000 rpm), ▲ 4.17 mg/mL (14,000 rpm), ▼ 4.17 mg/mL (15,000 rpm). The upper graph shows the fitting residuals of the lower sedimentation curves to a monomer-dimer model; the random, nonsystematic distribution of the residuals indicates a good fit of the data.

Mx is the molar extinction coefficient, $\wp$ is the path length of the centerpiece in cm, and *n* is the stoichiometry of the larger association species.

$$K_{conc} = K_{fringe} \times (dn/dc \times \wp / \oplus)^{n-1} \times (M_1)^{n-1} / n$$

where $K_{conc}$ is the association constant in molar concentration terms, $K_{fringe}$ is the signal association constant, dn/dc is the specific refractive increment, $\wp$ is the path length of the centerpiece in cm, $\oplus$ is the lightsource wavelength in cm, $M_1$ is the monomer molecular weight, and *n* is the stoichiometry of the larger association species.

The Sednterp program ***(5)*** is employed to calculate the partial specific volume of the proteins from the amino acid compositions using the method of Cohn and Edsall ***(8)***. The program also calculates the solvent density using known values from physical tables.

## 4. Notes

1. Sample cell leakage is one of the more common problems encountered with the analytical ultracentrifuge. This can be remedied by careful assembly of sample cell components, the use of Spinkote grease on the cell housing gasket and screw ring, tightening of the cell components to recommended torque pressures, and avoiding the mixing of cell components from different sample cell sets. Care should be taken not to apply excessive amounts of Spinkote grease because it will end up being sprayed onto the walls of the vacuum chamber at high rotor speeds and possibly coat the lenses of the optical systems.
2. The rotor vacuum chamber should be wiped clean with a lint-free cloth between each sample run and the lenses checked and cleaned if necessary.
3. The flash lamp for the Absorbance Optical System gets a coating on it over time and intensity checks should be performed periodically as detailed in the XLI instruction manual and the lamp cleaned if necessary.
4. To reduce problems with systematic fringe distortion in the Interference Optical System, which will adversely affect the fitting of the data, blank scans should be performed at the same speeds as the sample scans using the same sample cells filled with just buffer solution. These blank scans should then be subtracted from the sample scans prior to fitting.

## References

1. (1990) *Instruction Manual for the Dawn Model F*, Wyatt Technology Corp., Santa Barbara, California.
2. (1994) *Wyatt Optilab 903 Interferometric Refractometer Instruction Manual and Software Guide*, Wyatt Technology Corp., Santa Barbara, California.
3. Wyatt, Philip J. (1993) Light scattering and the absolute characterization of macromolecules. *Anal. Chimi. Acta.* **272,** 1–40.
4. (1997) *Optima XL-I Analytical Ultracentrifuge Instruction Manual*, Spinco Business Center of Beckman Instruments, Palo Alto, California.

5. Hayes, D. B. (Magdelan College), Laue, T. (University of New Hampshire), and Philo, J. (Amgen) (1995–1998) *Sedimentation Interpretation Program*, Version 1.01.
6. Johnson, M. L., Correia, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) Analysis of data from the analytical ultracentrifuge by non-linear least-squares techniques. *Biophys. J.* **36,** 575–588.
7. Yphantis, David A. (1991) in *Nonlinear Least Squares Program for Analysis of Equilibrium Ultracentrifugation Experiments*, Mansfield Center, Connecticut.
8. Cohn, E. J. and Edsall, J. T. (1943) Chapter 4, in *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Rheinhold, New York, p. 157.

# 12

# Small-Angle Solution Scattering Reveals Information on Conformational Dynamics in Calcium-Binding Proteins and in their Interactions with Regulatory Targets

**Jill Trewhella and Joanna K. Krueger**

## 1. Introduction

When a beam of neutrons or X-rays encounters a protein in solution, a small portion of the beam will be deflected or "scattered." The angular dependence of this scattering is related to the structure of the protein. For a solution of randomly oriented proteins, the scattering is concentrated in the vicinity of the direct beam, or zero-angle. Solution scattering is therefore often referred to as small-angle or low-angle scattering. Structural information encoded in the scattering data includes the overall size and shape of the protein. Although this information is relatively low-resolution, it is not limited by the requirement of having crystals and it can be applied to structures with dimensions in the range 10–1000 Å. This range is extremely useful for studies of proteins and the complexes they form.

The calcium-binding proteins have proven ideal candidates for study using small-angle solution-scattering techniques. Small-angle scattering studies of the dumbbell-shaped calcium-binding proteins, calmodulin (CaM) and troponin C (TnC), have been particularly fruitful. They have provided key insights into the conformational dynamics of the multifunctional CaM in different stages of the $Ca^{2+}$ signaling sequence, as well as into the nature of different CaM target enzyme interactions (*see* **Subheading 3.5.1.**). In the case of TnC, small-angle neutron scattering experiments have provided important structural constraints regarding its interactions with troponin I (*see* **Subheading 3.5.2.**).

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

This chapter will outline the practical aspects of small-angle scattering using X-rays or neutrons, the underlying theory of the technique, and approaches to scattering data analysis. Examples of scattering experiments on calcium-binding proteins and their interactions with the proteins whose activities they regulate will be highlighted in order to illustrate the information that can be obtained. Comprehensive reference texts on the theory and practice of small-angle scattering that are extremely useful have been written by Glatter and Kratky ***(1)*** and Feigin and Svergun ***(2)***. There is also a recent review ***(3)*** that gives an overview of the techniques and their application to biomolecular structural analysis in general.

## 2. Materials

### *2.1. Small-Angle Scattering Facilities*

A small-angle scattering instrument generally consists of (1) a radiation source that provides a beam of neutrons or X-rays; (2) a monochromator that selects the wavelength of the radiation; (3) a collimator that directs or focuses the radiation to a point on a detector; (4) a sample environment between the source and the detector; and (5) a position sensitive detecting device. The detectors are one- or two-dimensional proportional counters. One-dimensional detectors are less expensive, but with a two-dimensional area detector, the data can be circularly averaged to give a one-dimensional scattering profile with maximum signal to noise. This feature is particularly useful at the larger scattering angles where the scattering signal is weakest. In general, small-angle scattering instruments for protein studies are found at large user facilities, or are custom built in specialist laboratories. Synchrotron and neutron facilities have user programs whereby scientists can apply for time on the instruments, usually via a peer reviewed proposal system. Interacting with an established small-angle scattering laboratory or a user facility is often the best way to get started.

#### *2.1.1. X-Ray Scattering Sources and Instruments*

X-rays are produced either at synchrotron sources, where the radiation is emitted by the acceleration of electrons circulating at ultrahigh speed, or by laboratory X-ray generating devices such as X-ray tubes or rotating anodes. Laboratory devices depend upon the electronic transitions in atoms that emit X-rays (e.g., the K$\alpha$ transition of copper which yields 1.54 Å X-rays). The useful X-ray wavelengths for small-angle scattering are approx 1–2 Å. Synchrotron radiation has very high brilliance (intensity) which facilitates rapid data acquisition on small samples. Laboratory devices are comparatively weak, but they cause less damage to samples due to the ionizing radiation. There are

small-angle scattering instruments available for general users at a number of the U.S. synchrotron facilities including the Stanford Synchrotron Radiation Laboratory, the Advanced Photon Source at Argonne, and the National Synchrotron Light Source at Brookhaven. In addition there are facilities in Japan (The Photon Factory, Tsukuba) and Europe (ERSF in Grenoble, SRS in Daresbury, HASYLAB in Hamburg, and LURE in Orsay).

### 2.1.2. Neutron Scattering Sources and Instruments

Two types of neutron sources exist today: steady-state sources where neutrons are generated at nuclear reactors by nuclear fission, and pulsed sources where neutrons are generated by spallation at accelerators. The neutrons produced are generally of very high energy (short wavelength) and have to be "cooled" using moderators. Thermal neutrons refer to neutrons whose kinetic energies are close to the average energy of particles in materials at ambient temperatures. For small-angle scattering, the neutrons are further cooled down by low-temperature moderators (e.g., liquid hydrogen) and are referred to as "cold" neutrons. The wavelengths of cold neutrons lie in the region approx 2–20 Å. At this time, the steady-state reactor neutron sources have the best small-angle neutron instruments for protein studies because they provide the highest cold neutron fluxes. High-flux small-angle neutron instruments most suitable for protein structural studies are available for general users at the National Institute of Standards Technology in Washington, DC, and at the Institut Laue-Langevin, Grenoble, France. There are plans for an additional facility that will be optimized for structural biology applications at the HFIR reactor at Oak Ridge National Laboratory, TN.

## 2.2. Sample Requirements and Deuteration of Proteins

Samples for X-ray and neutron scattering optimally should have transmission factors of approx 0.7. The transmission factor is simply the ratio of the number of unscattered neutrons or X-rays that pass through the sample over the number absorbed or scattered. Protein concentrations for scattering experiments are generally of the order of mg/mL. The higher the concentration, the stronger the scattering signal. However, aggregation and/or interparticle interference effects can give problems at higher concentrations (*see* **Subheading 3.3.**), and so the optimal concentration must be evaluated for every system. For X-rays, the sample cross sectional areas are typically $mm^2$ and the thickness for optimal transmission is approx 0.7 mm, giving sample volumes of 10–50 μL. Samples for neutron scattering are larger, with approx 1 $cm^2$ cross-sectional area. The thickness of neutron scattering samples varies from one to several millimeters, depending on the hydrogen content of the solvent. Hydrogen has a very large incoherent neutron cross-section that contributes to the background

of the scattering signal and also results in multiple scattering (neutrons scattered more than once in the sample). Hence, the more hydrogen there is in a sample, the thinner it has to be so as to minimize these effects. A good rule of thumb is to never have more than 1-mm equivalent thickness of $H_2O$ in your sample.

Neutron-contrast variation experiments on protein complexes require the preparation of deuterated proteins. This deuteration cannot be achieved by simple exchange methods because it must include the nonexchangeable hydrogens in the protein. The level of deuteration needed depends upon the relative sizes of the proteins in the complex to be studied (*see* **Subheading 3.1.2.**). Deuterated proteins are prepared by growing *Escherichia coli* expression systems on deuterated substrates (i.e., $D_2O$ with a deuterated carbon source). The neutron experiment requires the deuteration to be as uniform as possible in order to minimize scattering density fluctuations within the deuterated component. High levels of deuterium in the growth medium slow the growth rate of *E. coli* and can dramatically reduce protein yields, and hence, rich media are often needed. Deuterated algal hydrolysate in $D_2O$ provides an excellent rich medium for *E. coli* expression systems. A minimal $D_2O$ media with deuterated glucose, glycerol, or sucrose will give sufficient yields for some systems. Growing on $D_2O$ with nondeuterated sugars results in uneven labeling that give rise to scattering density fluctuations that can be troublesome when analyzing the scattering data. To obtain less than 100% uniformly labeled proteins deuterated and nondeuterated nutrients are mixed proportionately in the growth medium and the incorporation must be tested (**Subheading 3.1.2.**).

### 2.3. Sample Cells and Holders

Sample cells must have windows that can contain the sample, but are as "transparent" to the probe radiation as possible. Commercially available thin-walled glass capillaries (1-mm diameter) make good X-ray sample containers. Although variability in the wall thickness and their round shapes mean that there must be a mechanism to position the capillaries very precisely and reproducibly in order to be able to accurately subtract sample and background measurements during data reduction (*see* **Subheading 3.2.**). Alternatively, X-ray sample cells can be designed using other materials (plastic, ceramics, metals) with windows made of thin (approx 0.01 mm) mica or beryllium.

Sample containers for neutron experiments can be conveniently made using quartz. Many neutron facilities design their sample holders so they can hold the typical 1-$cm^2$ quartz cell used in UV/visible spectrometers. These cells are available with precise sample path lengths in the range 1–5 mm, and their cross-sectional area is convenient for typical neutron beam sizes (approx $cm^2$). Circular quartz cells suitable for neutron experiments can also be purchased commercially.

## 3. Method

### 3.1. Underlying Theory and Scattering Data Interpretation

#### 3.1.1. The Basic Scattering Equation

X-rays and neutrons have the properties of plane waves, and the scattering pattern at small angles results from the constructive interference of secondary waves that are generated when a plane wave interacts with matter. Interference occurs when the secondary waves are additive. This additive property requires the scattering to be elastic (i.e., no energy change between the incident and scattered wave) and coherent (i.e., there is a defined phase relationship between the incident and scattered wave). The scattering profile for a macromolecule, such as a protein, in solution is a maximum at zero-scattering angle and falls off with a rate that depends upon the protein's size and shape; the larger the protein, the faster the fall off. The coherent, elastic scattering $I(Q)$ from a homogeneous solution of monodisperse proteins can be expressed mathematically as

$$I(Q) \propto \langle \int [\rho(r) - \rho_s] e^{-iQ \cdot r} dr|^2 \rangle \quad (1)$$

The integration is taken over the volume of the protein, and $\langle\ \rangle$ denotes the average over all orientations. $Q$ is the momentum transfer or scattering vector and its amplitude is $4\pi(sin\theta)/\lambda$, where $\theta$ is half the scattering angle and $\lambda$ is the wavelength of the scattered radiation. $\rho(r)$ and $\rho_s$ are the scattering densities for the protein and its solvent, respectively, and the difference between them is the "contrast." Scattering densities are calculated by summing the scattering amplitudes of each atom within a volume and dividing by that volume. The intensity of the scattering signal is proportional to the square of the molecular weight of the scattering molecules and to their number density (i.e., protein concentration). For a solution containing a mixture of different structures, the measured scattering will be the average of all the structures present weighted by their relative concentrations and the square of their molecular weights.

#### 3.1.2. X-Ray vs Neutron Scattering

X-ray sources, even the simple laboratory devices, are many orders of magnitude stronger than neutron sources. Because biological molecules are weak scatterers the relatively low fluxes of neutron sources is a disadvantage. However, neutrons are nonionizing radiation and hence quite benign. More importantly they offer the opportunity for "contrast variation" studies. X-rays are scattered by electrons and hence X-ray scattering amplitudes of atoms increase monotonically with the number of electrons. Neutrons are scattered by atomic nuclei, and neutron scattering amplitudes depend upon the complex properties of the neutron-nucleus interaction showing no systematic dependence on atomic number. Importantly, isotopes of the same element can have very differ-

**Table 1**
**Coherent Neutron Scattering, $b_{coh}$, and X-Ray Scattering, $f_{X\text{-}ray}$, Amplitudes for Biologically Relevant Nuclei**

| Atom | Nucleus | $b_{coh}$ ($10^{-12}$ cm) | $f_{X\text{-}ray}$ for θ = 0 in electrons (and in units of $10^{-12}$ cm) |
|---|---|---|---|
| Hydrogen | $^{1}H$ | – 0.3742 | 1.000 (0.28) |
| Deuterium | $^{2}H$ | 0.6671 | 1.000 (0.28) |
| Carbon | $^{12}C$ | 0.6651 | 6.000 (1.69) |
| Nitrogen | $^{14}N$ | 0.940 | 7.000 (1.97) |
| Oxygen | $^{16}O$ | 0.5804 | 8.000 (2.25) |
| Phosphorous | $^{31}P$ | 0.517 | 15.000 (4.23) |
| Sulfur | Mostly $^{32}S$ | 0.2847 | 16.000 (4.5) |

ent neutron-scattering properties. For neutrons, one of the largest differences is between the isotopes of hydrogen ($^{1}H$ = hydrogen, and $^{2}H$ = deuterium). **Table 1** gives the coherent, elastic X-ray, and neutron-scattering amplitudes for the atoms commonly found in biological systems. Note that the neutron scattering amplitudes for most nuclei, including $^{2}H$, are positive and approximately equal. The exception is the negative scattering amplitude for $^{1}H$ that results from a 180° phase shift between neutrons scattered by $^{1}H$ compared to the other nuclei. As a consequence, the neutron scattering density of a molecule depends strongly on its mean hydrogen content, and deuterium substitution can be used to manipulate neutron scattering densities.

### *3.1.2. Neutron Scattering and Contrast Variation*

**Equation 1** shows that the intensity of the scattering from a protein in solution depends upon the difference in scattering density between the particle and the solvent, i.e., its "contrast." If a complex is made using one deuterated and one nondeuterated protein, the two proteins will have very different neutron scattering densities. Further, by changing the deuterium level in the solvent, the neutron scattering contrast of each protein can be systematically varied. By adjusting the deuteration level in the solvent so that the mean solvent density matches that of either the deuterated or the nondeuterated protein, it is possible to "solvent match" that protein such that it has zero contrast and becomes "invisible" in the neutron experiment. Solvent matching thus provides a means for extracting structural information on the individual components within a complex.

A more robust approach to utilizing contrast variation methods with neutron scattering for extracting structural information on the components within the macromolecular complexes is to measure a "contrast series" in which the solvent deuteration level is systematically varied over he widest range possible.

For a complex of two components with different mean neutron-scattering densities, the total scattering can be written as

$$I(Q,\Delta\rho_A, \Delta\rho_B) = \Delta\rho_A{}^2\,(Q) + \Delta\rho_A\Delta\rho_B I_A(Q) + \Delta\rho_B{}^2 I_B(Q) \qquad (2)$$

The subscripts A and B refer to each component and $\Delta\rho_x = \rho_x - \rho_s$ where $\rho_x$ is the mean-scattering density for the individual components, i.e., $X = A$ or $B$. **Equation 2** assumes the difference between the mean-scattering densities for the individual components is much greater than any internal density fluctuations within each component. The three $I(Q)$ functions in equation 2 correspond to the three basic scattering functions. $I_A(Q)$ and $I_B(Q)$ represent the scattering of components $A$ and $B$, respectively, from which one can derive the structural parameters for each component within the complex. $I_{AB}(Q)$ is the cross term that yields information on the relative dispositions of the two components. A set of neutron scattering measurements with different $D_2O$:$H_2O$ ratios in the solvent gives a set of equations in the form of **Eq. 2** which can be solved using multiple linear regression to give the three basic scattering functions.

In any contrast variation experiment, it is crucial to know the precise level of deuteration in your sample. Without this information, you cannot calculate the contrast factors in **Eq. 2** that are required to solve for the basic scattering equations. For a complex of a deuterated and a nondeuterated protein, a plot of the square root of the forward scattering ($I_0$), normalized to molar protein concentration, as a function of the solvent scattering density (i.e., $D_2O$ content) will be linear and will cross zero at the scattering-density value corresponding to the solvent match point for the complex. From this value, and using the known chemical composition and partial specific volumes of the proteins in the complex, as well as knowledge of the number of exchangeable hydrogens, one can calculate the deuteration level. Alternative methods for estimating deuteration levels are to use NMR or mass spectrometry. To obtain good structural data on both components within the complex from a contrast series, it is best to measure scattering data on both sides of the solvent match point for the complex. As a result, it is important to choose the deuteration level so that the match point of the complex is around 50% $D_2O$.

### *3.2. Scattering Data Acquisition and Reduction*

In a scattering experiment, the protein solutions are placed in the sample cell and positioned in the X-ray or neutron beam to be irradiated while the intensity of the scattered radiation is measured as a function of angle at the detector. The X-ray or neutron beam must be smaller than the sample so that the entire beam passes through the sample. The detector must be calibrated for flatness of response and for precise $Q$ determination at each point on the detector. The detector response can be determined using an isotropic scatterer placed in the

beam, or alternatively using a radiation source that can uniformly flood the detector. For X-rays, we often use an $Fe^{55}$ radiation source to flood the detector. In the case of neutrons, the incoherent scattering from a 1-mm thickness of $H_2O$ provides a strong uniform scatterer for detector flatness calibration. The calibration of the $Q$-scale on the detector can be made using a sample with a known diffraction peak. Microcrystals of cholesteryl myristate, formed by melting and rapid cooling, give a relatively strong X-ray diffraction peak at $Q = 0.2483$ Å$^{-1}$ and a weaker one at half that value. These peaks can be used to calibrate the $Q$-scale on the detector and test for linearity in $Q$. Scattering facilities usually have a range of standards that allow for calibration of $Q$ scale and detector flatness.

The net scattering from protein molecules in a solution is calculated by subtracting the scattering measured from the solvent, including any buffer or salt present in the sample. The protein solution and solvent scattering data must be scaled to the same number of X-rays or neutrons incident on the sample for each measurement. Using the partial specific volume for the protein, one can then adjust the solvent subtraction for the small volume occupied by the protein molecules in the solution, i.e.,

$$I(Q)_{\text{protein}} = I(Q)_{\text{protein+solvent}} - (1 - V_p)\, I(Q)_{\text{solvent}} \tag{3}$$

where $V_p$ is the volume fraction occupied by the protein molecules in the solution.

The analysis of small-angle scattering data assumes that the source of radiation is a point, and that the wavelength of the radiation has a single value. In practice, the source has finite dimensions and a distribution of wavelengths. The beam-shape profile and wavelength distribution therefore may require a correction. At neutron sources, where the experiments are almost always intensity limited, the instrument can be set to use a wide range of wavelengths in order to gain intensity without serious loss of information. The wavelength spread is generally given in terms of the full width at half-maximum of the wavelength distribution divided by the value of the wavelength for which the distribution is a maximum (i.e., $\Delta\lambda/\lambda$). These values can range from 0.10–0.35 depending on the size of the object being studied and the desired balance between resolution and intensity. Beam-shape profile and wavelength distribution correction algorithms are included in most of the scattering data analysis software that is available to users.

### 3.3. Concentration Effects

Analysis of scattering data for accurate determination of protein structural parameters requires the data be free of concentration-dependent effects arising from aggregation and interparticle interference. Interparticle interference arises

when electrostatic interactions between protein molecules give rise to a distance of closest approach such that their distribution in solution is correlated. These correlations give rise to an interparticle contribution to the scattering that is convoluted with the intraparticle contribution that contains the shape information ***(4)***. The interparticle contribution rises from 0, peaks at a value related to the distance of closest approach, and then oscillates about 1. The effect on the scattering data is to suppress the intensity profile at the lowest scattering angles. Interparticle interference effects generally increase linearly with concentration and can be eliminated by measuring scattering data at 5–10 different protein concentrations and extrapolating the data to infinite dilution (zero concentration).

### 3.4. Data Analysis and Interpretation

#### 3.4.1. Guinier Analysis for Determination of Radius of Gyration and Forward Scattering

Two extremely useful parameters that can be determined from small-angle scattering data are the radius of gyration $R_g$ and forward or zero-angle scattering $I_0$. $R_g$ gives a simple geometric measure of how extended the particle is. It is defined as the root-mean-square distance of all elemental scattering volumes from their center of mass, weighted by their scattering densities. $I_0$ is proportional to the square of the molecular weight of the scattering particle, its contrast, and to the molar protein concentration. If the protein concentration is converted to mg/mL, the proportionality is to the molecular weight instead of its square ***(5)***.

For X-ray experiments, most proteins in aqueous solution have essentially the same contrast. Therefore, by using a standard protein of known concentration that is also known to be monodisperse in solution, its $I_0$ value can be compared to that of the protein of interest with known concentration to test for monodispersity. For a mixture of two or more different proteins of known relative concentrations, complex formation can be monitored. In the case of a single protein species in solution that is known to be monodisperse, $I_0$ can be used to calculate very precise concentration values. The relationship used for these types of analyses, for protein concentrations given in mg/mL, is:

$$I_{0\,x}/M_x c_x = I_{0\,\text{standard}}/M_{\text{standard}} c_{\text{standard}} \tag{4}$$

where $x$ indicates the "unknown" protein, $c$ is the protein concentration, and $M$ the molecular weight of the protein. For studying complex formation involving two different proteins, it is more convenient to use molar protein concentrations and substitute $M^2$ for $M$ in **Eq. 4**.

At very low scattering angles, Guinier ***(6)*** showed that the scattering data could be conveniently approximated by an exponential function of the square of the momentum transfer

$$\ln I(Q) \cong \ln I_0 - Q^2R_g^2 / 3 \tag{5}$$

By plotting the logarithm of $I(Q)$ vs $Q^2$ estimates of $R_g$ and $I_0$ can be determined directly from the slope and intercept values, respectively. For a symmetric, globular shaped proteins, or protein complexes, the Guinier approximation is good for values of $Q$ for which $QR_g \leq 1.3$. For more asymmetric shapes, this range is reduced. In cases where one dimension of a protein, or assembly of proteins, is much greater than the other two (e.g., a rod) the scattering data can be scaled by $Q$ to remove the longer distance and the corresponding Guinier approximation becomes

$$\ln QI(Q) \cong \ln QI_0 - Q^2R_c^2 / 2 \tag{6}$$

where $R_c$ is the average radius of gyration of cross-section of the shape.

### *3.4.2.* P*(r) Analysis*

A very useful function for interpreting small-angle scattering data is the pair-distance, or vector-length distribution function $P(r)$. It is obtained by calculating the inverse Fourier transformation of the intensity profile $I(Q)$:

$$P(r) = (1 / 2\pi^2) \int I(Q)Q \cdot r \sin(Q \cdot r)dQ \tag{7}$$

$P(r)$ is simply the frequency of vector lengths connecting small-volume elements within the entire volume of the scattering particle weighted by the product of the scattering contrast at each element. $P(r)$ goes to zero at a value corresponding to the maximum dimension of the particle $d_{\max}$. $P(r)$ is more readily interpreted in terms of structural information than is the scattering profile $I(Q)$ because it is represents the real space distance information. It is extremely sensitive to the overall shape of the scattering particle, and to the relationships between domains or repeating structures. The $P(r)$ function can be used to calculate values for $R_g$ and $I_0$ that are more precise than the values obtained from Guinier analysis, because the entire scattering profile is used. The zeroth moment of $P(r)$ gives the forward scattering $I_0$ whereas $R_g$ is calculated as the second moment of $P(r)$:

$$R_g^2 = \int P(r)r^2d^3r / 2\int P(r)d^3r \tag{8}$$

The Fourier transform of the cross term $I_{AB}(Q)$ defined in **Eq. 2** (*see* **Subheading 3.2.1.**) yields a $P(r)$ function representing the vector length distribution function for all the vectors between components A and B in the complex.

This $P(r)$ function gives information on the disposition of the two components. For two interacting globular proteins the first moment of this $P(r)$:

$$\int P(r) r d^3r \,/\, \int P(r) d^3r \tag{9}$$

gives an estimate of the distance between the centers of mass of the two components.

Because $I(Q)$ can be measured only over a finite range of $Q$-values, $P(r)$ functions are generally calculated using an indirect Fourier transform method in which the coefficients of a set of functions in real space are optimized to fit to the scattering data. These coefficients are used to calculate the corresponding Fourier transform of the real space series in order to obtain the related scattering profile. Several of these methods have been developed using different basis sets ***(7,8)***. Alternatively a regularization method can be used ***(9)***. The available software packages using these methods generally include algorithms that can calculate the effects of beam shape and wavelength distribution on the scattering profile (*see* **Subheading 3.2.**)

### *3.4.3. Structural Modeling*

Although the type of information contained in parameters such as such as $R_g$, $I_0$, and $P(r)$ is sufficient for many experiments, modeling procedures are essential for interpreting the scattering data in more detail. Modeling studies can generally be divided into two categories. In cases where high-resolution structural data are available, they can be used as the basis for interpreting protein conformational changes under various experimental conditions. A model $P(r)$ function can be calculated based on the crystal structure and compared directly with that determined from solution scattering data ***(10)***. The model $P(r)$ function is most efficiently calculated using a rapid Monte Carlo integration routine in which the crystal structure coordinates are placed in a box and volumes are assigned either to each atom in the structure or to amino acid groups. The box is then filled with random points, and points that fall within the molecular boundary are saved along with a weight equal to the scattering contrast at that point. The $P(r)$ function is calculated simply by summing the distances between all saved pairs of points, weighted by the product of the respective scattering densities.

In the absence of high-resolution structural data, low-resolution models built up from uniform-density shapes can be used to aid in the interpretation of the scattering data ***(11,12)***. Such an approach is advantageous for conducting an initial search of conformational space with relatively few parameters to vary. Because the scattering profiles are generally not available for these objects

analytically, the same type of Monte Carlo integration techniques are used to calculate model $P(r)$ functions as for the crystal structures. High-resolution structures can subsequently be fit into the uniform density shapes for more detailed interpretations. These modeling approaches are particularly useful when combined with neutron contrast variation data and can provide information on the quaternary structure and interactions between components in multisubunit enzymes or interacting proteins (*see* **Subheading 3.5.**).

## 3.5. Illustrative Examples of Small-Angle Scattering Studies of Calcium-Binding Proteins

### 3.5.1. Calmodulin–Kinase Interactions

A number of the enzymes regulated by CaM are kinases, and the $Ca^{2+}$/CaM-dependent activation of myosin light-chain kinase (MLCK) serves as a model for CaM–kinase interactions. MLCK has a catalytic core that is homologous to that of other protein kinases ***(13)*** with a large and small lobe between which lies the catalytic cleft. This catalytic core is followed immediately by a carboxyl-terminal regulatory segment consisting of autoinhibitory and CaM-binding sequences. In its inhibited conformation, the regulatory segment of MLCK maintains numerous contacts with the catalytic core ***(14)***. The autoinhibitory hypothesis for $Ca^{2+}$/CaM regulation of MLCK proposes that CaM-binding to MLCK releases these contacts and opens the catalytic site for substrate binding and modification. CaM has an unusual dumbbell-shaped structure with two lobes, each containing two EF-hand motifs that bind $Ca^{2+}$ with similar affinities, connected by an extended helical region ***(15)***. When $Ca^{2+}$ binds to CaM, hydrophobic clefts on each lobe are exposed that are important in recognizing and binding to the CaM-binding sequences in its targets (reviewed in **ref. *16***).

Small-angle scattering studies of CaM in solution first detected the small extension of its dumbbell-shaped structure because of the opening of the hydrophobic clefts in each globular domain upon $Ca^{2+}$-binding ***(17)***, and subsequently provided structural evidence that the interconnecting helix region was flexible in solution ***(10)***. Soon after, small-angle scattering studies of CaM complexed with peptides that model the CaM-binding domains in target enzymes showed CaM undergoes a dramatic conformational change facilitated by the flexible interconnecting helix region (reviewed in **ref. *18***). **Figure 1** shows the $P(r)$ functions measured for CaM free in solution and in its complex with the peptide sequence (MLCK-I) corresponding to its binding domain in skeletal muscle MLCK. For comparison, the $P(r)$ function calculated from the crystal structure of $4Ca^{2+}$/CaM is also shown. The crystal structure based $P(r)$ function is characteristic of a dumbbell-shaped structure showing two distinct peaks; the first dominated by the most frequently occurring vector lengths

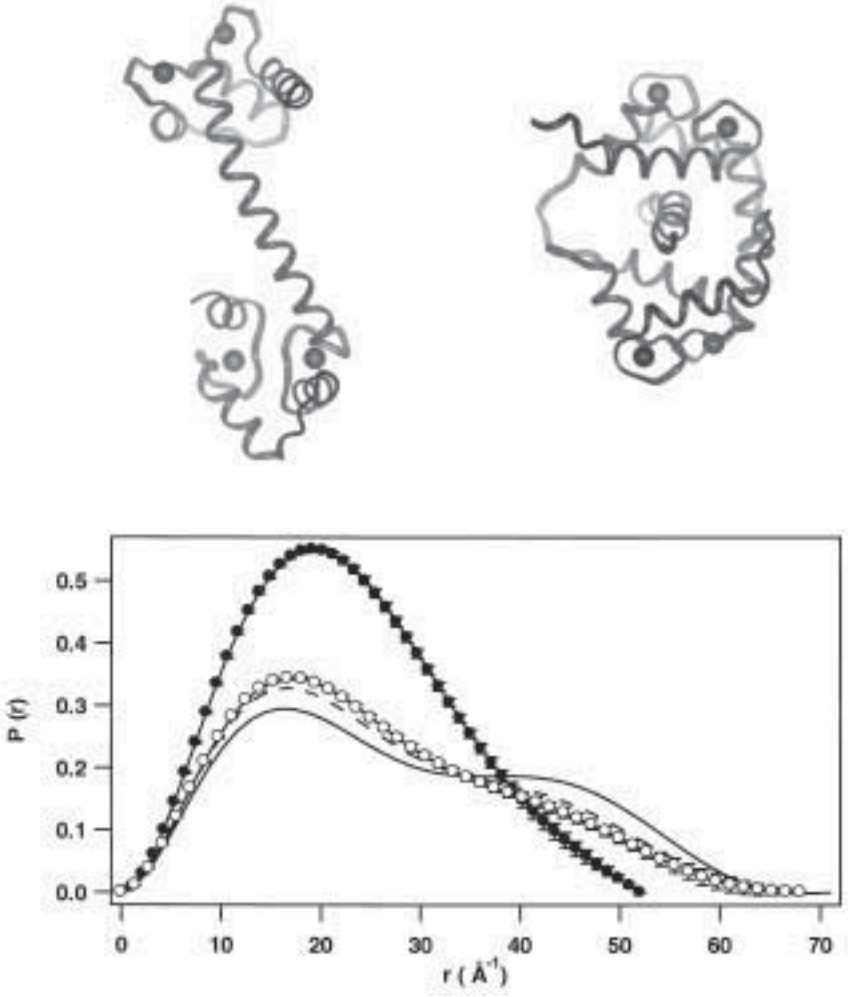


Fig. 1. Ribbon representation of the backbone structure of CaM in the crystal structure (*top left*, ***16***) and in its complex with the peptide MLCK-I derived from NMR data (*top right*, ***19***). The plots show the $P(r)$ functions, each scaled to the square of the molecular weight, calculated from the crystal structure of $4Ca^{2+}$/CaM (*solid line*), and measured using solution scattering from CaM (*dashed line*), $4Ca^{2+}$/CaM (*open circles*), and $4Ca^{2+}$/CaM/MLCK-I (*closed circles*).

within each of the individual globular lobes and the second dominated by vector lengths between the globular lobes. The experimental $P(r)$ measured for $4Ca^{2+}$/CaM in solution shows a reduction in the number of longer vector lengths compared with that calculated for the crystal structure, and the second peak has

collapsed to a shoulder at a shorter average vector length. These changes indicate that the two lobes, on average, are closer together in solution and are reflected in the $R_g$ values that go from 22.8 Å for the crystal structure to 21.3 Å for the solution form. The apo-CaM $P(r)$ is only slightly more compact than the $Ca^{2+}$-saturated form, giving an $R_g$ value of 19.6 Å. The CaM/MLCK-I complex has a symmetric $P(r)$ with a single peak and gives an $R_g$ value of 16.4 Å indicating a much more compact, globular shape. The $d_{max}$ value is also smaller, by about 20 Å, indicating that the globular lobes of CaM in the complex must come into close contact with each other. These scattering data were interpreted as indicating that CaM wrapped itself around the MLCK-I peptide by using the flexibility in the interconnecting helix region. Subsequent NMR experiments ***(19)*** confirmed the conformational collapse observed in the scattering involves the globular calcium-binding domains of CaM folding around the MLCK-I peptide that forms an amphipathic helix. The NMR data further show the complex is stabilized via hydrophobic interaction between the peptide and the hydrophobic clefts exposed upon $Ca^{2+}$-binding to each globular domain of CaM. The scattering data have thus contributed to our view of the highly conserved, multifunctional CaM being designed for both specific binding to targets as well as diversity in the set of targets it must regulate. Many of the CaM target sequences are approx 20 residues in length and characteristically have a high propensity for forming an amphipathic helix, with large hydrophobic side chains spaced 12 residues apart (reviewed in **ref. *16***). There is, however, considerable sequence variability along the length of the peptide. The large hydrophobic side chains act as recognition and binding sites for the hydrophobic clefts, whereas the interconnecting helix region provides the flexibility to accommodate a variety of stereo-chemical surfaces.

A number of CaM/peptide studies indicate that the conformational collapse observed in the CaM/MLCK-I example is a common feature of a number of $Ca^{2+}$–CaM activation mechanisms (reviewed in **refs. *16*** and ***18***). However, there are no high-resolution structures of CaM complexed with any one of the more than 30 enzymes it is known to regulate, leaving a number of questions about the applicability of the results of the CaM/peptide studies to the enzyme activation mechanism.

Would the extensive network of interactions between the CaM-binding and autoinhibitory sequences of MLCK inhibit the formation of the compact structure seen in the CaM/peptide complexes? Would the presence of other structural elements of the enzyme sterically hinder its formation? Small-angle neutron contrast variation studies answered these questions definitively and revealed the first structural view of CaM complexed with a catalytically active enzyme, MLCK ***(11,12)***. Structural models for the MLCK/deuterated-CaM complex with and without bound substrates were obtained (*see* **Fig. 2**). The

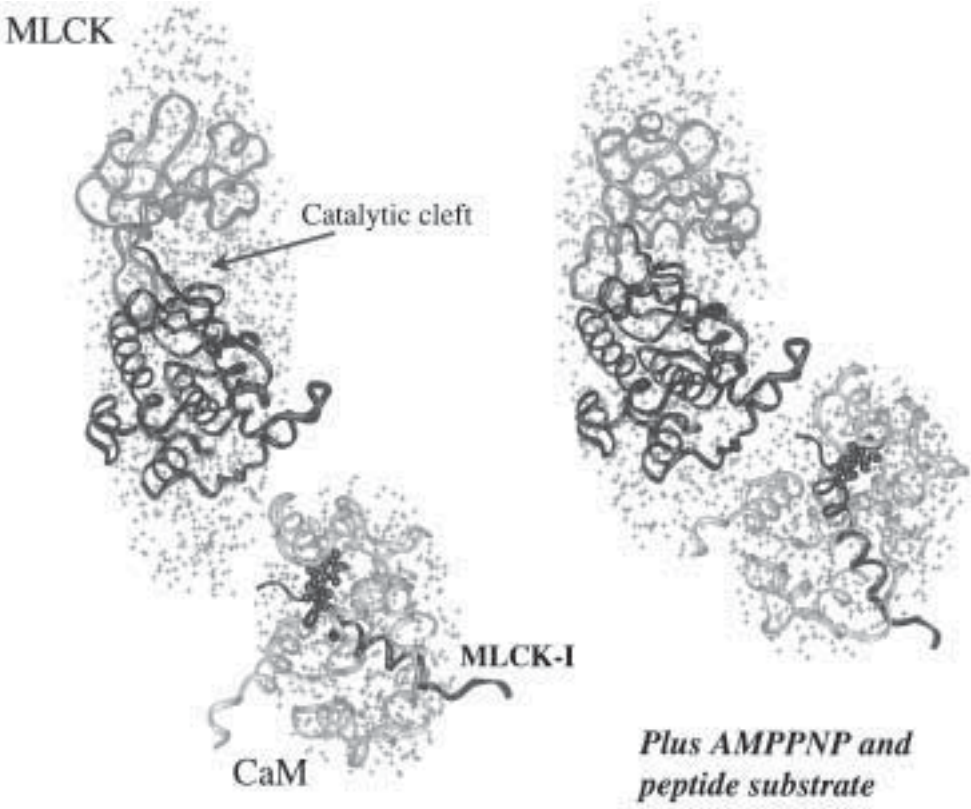


Fig. 2. Models derived from the neutron scattering data for the $4Ca^{2+}$/CaM/MLCK complexes with (right, ***12***) and without (left, ***11***) substrates. The conserved portion of the inhibited kinase catalytic core ***(13)*** and the NMR structure of CaM complexed with MLCK-I ***(19)*** are fit within the dimensions of the larger and the smaller ellipsoids, respectively, that were derived from modeling the neutron scattering data. The upper and lower lobes of the catalytic core are represented as gray and black ribbon drawings, respectively, with the catalytic cleft between them labeled. The empty spaces in the ellipsoid representing MLCK are occupied by *N*- and *C*-terminal sequence segments whose structures are not known at this time. CaM is represented as a gray ribbon drawing, with its bound MLCK-I peptide in black, and a CPK representation of its hydrophobic Trp residue near the *N*-terminal end. This Trp residue is key to recognition and binding by the *C*-terminal CaM domain. The models show that CaM binds to the kinase such that there must be a significant movement of the CaM-binding and autoinhibitory sequences away from the surface of the catalytic core. Upon substrate binding there is a movement of the CaM approx 12 Å closer to the catalytic cleft accompanied by a reorientation that brings the *N*-terminal helix of CaM into contact with the surface of the kinase. At the same time the catalytic cleft of MLCK closes, presumably about its substrate. This contraction is inferred from the overall shortening of the MLCK ellipsoid seen in the neutron model derived from the plus substrate experiment. Figure adapted from **ref. *12***.

$P(r)$ function derived for CaM within the complex shows that it undergoes an unhindered conformational collapse upon binding MLCK that is indistinguishable from that observed with the isolated CaM-binding peptides. This result requires that the CaM-binding domain, as well as some portion of its neighboring residues in the sequence be completely removed from interactions with the catalytic core. The model further shows that CaM binds to the enzyme at a site that is distant from the catalytic cleft, which also requires a significant movement of the autoinhibitory sequence away from the surface of the catalytic core. These data provided the first direct structural evidence for the autoinhibitory hypothesis for MLCK activation.

The neutron experiments further revealed that the binding of a peptide substrate and a nonhydrolyzable analog of ATP (AMP.PNP) to the $4Ca^{2+}$/CaM-bound kinase results in a "closure" of the enzyme's catalytic cleft, thus bringing together all the elements required for catalysis. In addition, the separation of the centers-of-mass of the CaM and MLCK components is shortened (by approx 12 Å bringing CaM closer to the catalytic site. Finally, there is a reorientation of CaM with respect to the kinase upon substrate binding that results in interactions between the *N*-terminal sequence of CaM and the kinase that were not observed in the complex without substrates. This reorientation is of particular interest in light of the observation that deletion of the *N*-terminal leader sequence of CaM abolished CaM-dependent activation of skeletal muscle MLCK although having no effect on the apparent affinity ***(20)***. The neutron-derived models thus provide evidence that there is an interaction between the *N*-terminal helix of CaM and the surface of the kinase that is important for activation.

Small-angle X-ray scattering has recently provided evidence for a $2Ca^{2+}$/CaM/MLCK intermediate in the MLCK activation mechanism ***(21)***. Analysis of the forward scattering, or $I_0$, data from a $Ca^{2+}$ titration of CaM/MLCK complexes showed that there is full complex formation when there is only two mole equivalents of $Ca^{2+}$ per complex. This $2Ca^{2+}$ intermediate had been suggested to exist under physiological conditions in the absence of a $Ca^{2+}$ signal ***(22,23)***. The purpose of such an intermediate could be to restrain the CaM from diffusing away in the absence of the $Ca^{2+}$ signal in rapidly cycling functions such as muscle contraction and relaxation. Because the $Ca^{2+}$ affinities of CaM are strongly affected by its different target binding sequences, it has been further suggested that CaM-binding sequences in different enzymes may serve the purpose of "tuning" the calcium affinities of the $Ca^{2+}$-binding sites so as to optimize for the formation of such intermediates when needed. Thus, in the CaM/MLCK example, we see the *C*-terminal domain of CaM may in fact be functioning as an "anchor" whereas CaM's *N*-terminal lobe would possess the regulatory function, alternately binding and releasing the autoinhibitory sequence of MLCK in response to the $Ca^{2+}$ signal.

The CaM/MLCK complexes are too large for NMR structural analysis, and the results of the scattering experiments suggest that they may continue to resist study by crystallography because they are inherently flexible. The scattering data thus provide a critical framework in which the high-resolution structural data for the individual components can be placed in order to understand details of their interactions. **Figure 3** summarizes this framework derived from small-angle scattering experiments that has contributed to our current understanding of the CaM/MLCK activation mechanism.

Although the CaM/MLCK system is broadly used as a model for CaM/enzyme activation mechanisms, small-angle scattering studies have shown that there are CaM/target enzyme interactions that have quite distinctive features. For example, CaM is an integral part of the multisubunit enzyme phosphorylase kinase whose CaM-binding domain consists of two subdomains approx 25 resides each. Small-angle scattering experiments have shown that CaM remains extended when bound to both these subdomains ***(24)***. This different type of interaction may be important in maintaining the $Ca^{2+}$-independent association of CaM with the other subunits in phosphorylase kinase that may be characteristic of this more complex multisubunit enzyme. Small-angle scattering thus promises to continue to be an extremely useful technique for probing the conformationally dynamic and structurally diverse CaM/target enzyme complexes.

### *3.5.2. Troponin C–Troponin I Interactions*

Like CaM in phosphorylase kinase, TnC is an integral component of a multisubunit complex. Troponin (Tn) has three components designated C, I, and T. TnC regulates the interactions of the inhibitory TnI with the thin filament in muscle in a $Ca^{2+}$-dependent manner thereby controlling thick and thin-filament interactions. TnC is structurally and functionally homologous to CaM, with an important distinctive feature of having two orders of magnitude difference in the $Ca^{2+}$ affinities between the two globular domains. The *C*-terminal domain has the high-affinity binding sites that are believed to be always occupied in muscle, whereas the *N*-terminal domain has the lower affinity, $Ca^{2+}$-specific regulatory sites. Neutron-scattering experiments using contrast variation on complexes of TnI with deuterated TnC (skeletal forms) in solution ***(25)*** have provided structural data on the overall complex and its components. The $P(r)$ function for TnC within the complex is very similar to that calculated using the crystal structure, indicating a fully extended interconnecting helix region, perhaps reflecting a structural and functional similarity to the CaM/phosphorylase kinase interactions ***(24)***. The neutron experiments also showed that TnI within the TnC–TnI complex has an even more extended structure than TnC, with a maximum linear dimension that is the same as for the overall complex. More recently, Stone et al. ***(26)*** published the results of solvent matching neutron scattering experiments on reconstituted troponin containing either

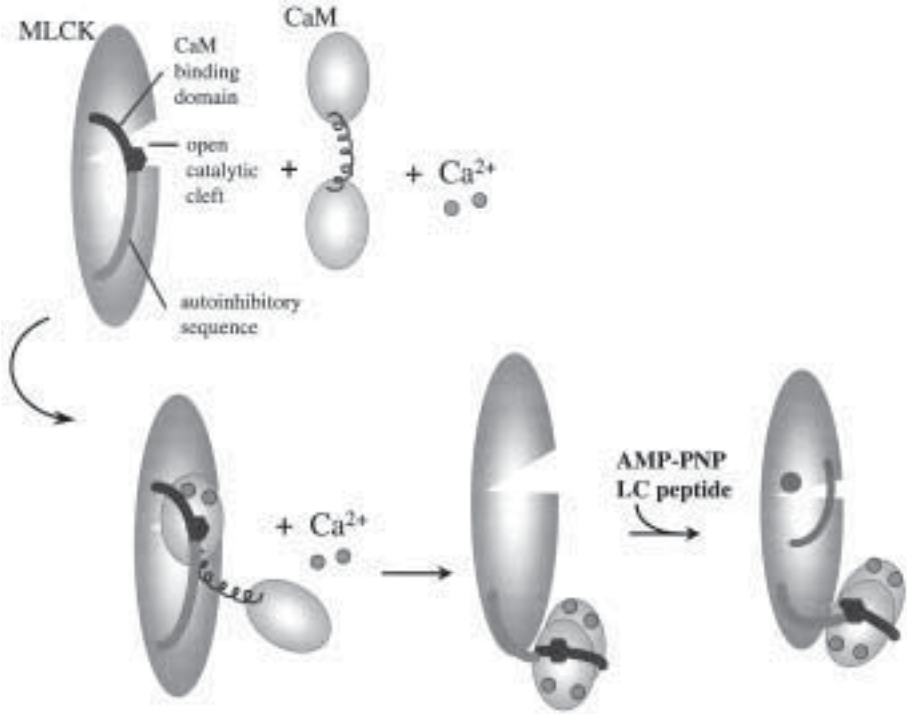


Fig. 3. Schematic summarizing the sequential conformational transitions for CaM activation of MLCK determined from small-angle scattering. In its inactive conformation, the regulatory segment of MLCK maintains numerous contacts with the catalytic core and apo-CaM is shown with a flexible interconnecting helix. Upon binding $2Ca^{2+}$, the hydrophobic cleft of the *C*-terminal lobe of CaM is opened and a $2Ca^{2+}$–CaM–MLCK complex forms ***(21)*** via an interaction with the Trp residue in the CaM-binding sequence. (This interaction may be maintained in vivo even in the absence of the $Ca^{2+}$ signal; *see* text.) Addition of another $2Ca^{2+}$ opens the hydrophobic cleft in the *N*-terminal lobe of CaM that binds to a hydrophobic group 12 residues away from the initial Trp recognition residue. This second interaction with the *N*-terminal lobe drives the conformational collapse of CaM. MLCK autoinhibition is relieved as this binding induces a significant movement of the MLCK regulatory segment away from the surface of the catalytic core ***(11)***. Once CaM releases the autoinhibition of MLCK, substrate binding induces closure of the kinase catalytic cleft as well as a movement of the CaM center-of-mass toward that of MLCK ***(12)***. At the same time, CaM reorients with respect to the kinase so as to bring about a close interaction between the *N*-terminal leader sequence of CaM and the kinase. Thus the fully activated complex is formed. Figure adapted from **ref. *12***.

deuterated TnI or deuterated TnC in the presence and absence of $Ca^{2+}$. These experiments confirm that TnC has the same radius of gyration within the troponin complex as in the crystal structure, and that TnI has a highly elongated

structure. They further show that TnI undergoes a compaction upon addition of $Ca^{2+}$. The measurements of the radius of gyration for TnI in the ternary troponin complex indicate a more compact structure than was observed for the binary complex, suggesting that of TnT influences the TnC–TnI interaction.

The TnC–TnI interaction has proven very difficult to study, and there is a growing amount of apparently contradictory data using different experimental approaches. One difficulty the system has for study using scattering techniques is that there are no high-resolution structural data on TnI and its structure is very unusual. TnI is not a "typical" globular protein, and therefore cannot be modeled simply as an ellipsoid to aid in the interpretation of scattering data. Based on the neutron data for the binary complex, a model was developed ***(27)*** in which TnI forms a superhelical structure around the $4Ca^{2+}$/TnC extending into an incomplete donut shaped structures that project beyond the TnC at each end. The diameter of the TnI central spiral is 12 Å, close to that expected for an $\alpha$ helix, and it passes through or near the two hydrophobic clefts in each globular domain of $4Ca^{2+}$/TnC. Based on this model, it was proposed that the *C*-terminal domain of TnC anchors TnI while the *N*-terminal domain alternately binds and releases TnI in response to the $Ca^{2+}$ signal. Recent data from NMR, crosslinking FTIR, and crystallography have yielded a high-resolution model for Tn(–TnI) that is based on the earlier neutron model ***(28)***. Thus, although there remains much work to be done before we fully understand the complex interactions within the troponin complex, the neutron scattering experiments on TnC–TnI ***(25)*** and troponin ***(26)*** have yielded some important consensus conclusions. In the presence of the *intact* TnI, TnC has a fully extended structure that is very similar to its crystal structure. Further, TnI has an even more elongated structure, although the details of that structure show a dependence on TnT and $Ca^{2+}$.

## 4. Notes

1. Protein aggregation and the importance of accurate protein concentration determinations and $I_0$ calibration: Nonspecific protein aggregation can be fatal in a small-angle scattering experiment. Analysis of scattering data from solutions containing aggregates will give structural parameters that are systematically larger than the correct values. Because the scattering signal is proportional to the square of the molecular weight of the scattering particle, even small amounts of aggregated material in a solution will bias the data very severely toward the aggregated species. If it is severe enough, aggregation can be seen in the very small angle scattering data as a deviation from linearity in the Guinier plots (*see* **Subheading 3.5.1.**). However, samples with small amounts of aggregation can yield perfectly linear Guinier plots. Another signature of aggregation can be seen in the concentration dependence of the scattering data. For a homogeneous, monodisperse solution of protein molecules, the shape of the scattering profile should be either

concentration independent or show evidence of suppression of the lowest-$Q$ data because of interparticle interference (*see* **Subheading 3.3.**). This suppression leads to increasingly underestimated structural parameters with increasing concentration. If the concentration dependence of the structural parameters goes in the opposite direction (i.e., increasing structural parameters with increasing concentration), then it is evidence for concentration-dependent aggregation, and extrapolation of the data to infinite dilution is unlikely to yield the true structural parameters of the monomeric protein.

The best guard against being misled by aggregation is to place the scattering data on an absolute scale in order to obtain an accurate measure of the molecular weight of the scattering particle ***(29)***. Alternatively, the forward scattering from the protein of interest can be calibrated using a standard protein known to be monodisperse. A good protein standard is one that has a molecular weight close to the value of the protein you wish to study, and whose concentration can be accurately determined. A protein with a well-determined UV extinction coefficient can be a good choice. Apoferritin has been used as a standard, as has lysozyme ***(5)***. However, many proteins are not suitable for use as a standard at synchrotron intensities because of radiation induced aggregation and standards must be reevaluated for each new instrument.

2. Hydration layer effects: Macromolecules in aqueous solution have water molecules at their surfaces that can form a "hydration layer" with a mean scattering density that is different from that of the bulk solvent. In terms of the scattering experiment, the protein solution then becomes a three component system (protein, hydration layer, bulk solvent; ***28***). The effect of the hydration layer is to give a scattering pattern the yields structural parameters that are larger than for the protein by itself. The size of the hydration layer effect depends upon the charge on the protein surface and the ionic strength of the solution. For highly charged molecules, like DNA for example, the effects are significant. For close to neutral proteins in dilute buffered salt solutions (<250 m$M$), these effects can be insignificant. Each new system needs to be evaluated. Because the contrast factors for the hydration layer are different for neutrons and X-rays, by combining X-ray and neutron scattering data on the same sample, it is possible to determine the thickness and contrast of the hydration.
3. Effects of ionizing radiation: Radiation damage from ionizing X-rays, especially at a synchrotron source, often results in protein aggregation. The production of free radicals in the protein solution by the X-rays can cause bond breakage in the protein molecules that can result in aggregation. These effects can be minimized by using antioxidants that can keep the free radical concentrations in check. At high-intensity synchrotron sources, it is common practice to check for radiation-induced aggregation by tracking the scattering profile as a function of time in intervals as short as 30–120 s. Aggregation is observed as a time-dependent increase in the intensity of the small-angle scattering at the lowest $Q$-values. One should routinely check for a protein's sensitivity to X-rays by doing gel electrophoresis to look for cleavage products in samples that have been X-rayed.

4. Neutron solvent matching vs contrast series measurement: The effectiveness of the solvent-matching experiment depends upon having uniform density components such that the internal density fluctuations can be ignored, as well as on very precise matching of the component and the solvent densities. Both of these conditions can be difficult to achieve, and failure will significantly bias the scattering data. In addition, a solvent matching experiment provides only one scattering profile against which to evaluate models. A neutron contrast series can provide as many scattering profiles as the experimenter can make samples with different contrasts. This larger data set can more rigorously test more complex models.
5. The perils of modeling: The interpretation of small-angle solution scattering data in terms of 3-D models is perilous because more than one 3-D structure can yield the same one-dimensional scattering profile! As a result, scattering data can unequivocally prove a model incorrect, but by themselves they cannot prove a model correct. The significance of a model that is developed based on comparisons with scattering done must be evaluated in terms of the assumptions and constraints that have been applied to the model search, as well as the number of variables used. With sufficient constraints on a system, it is possible to exhaustively search conformational space and find the solution, or set of very similar solutions, that best satisfy the scattering data within the constrained set tested. Statistical tests (e.g., the $\chi^2$ test) can be applied to assess how unique the best-fit solutions are. An example of a constrained model search would be to search all possible ellipsoid shapes with a volume consistent with the molecular weight and partial specific volume of a protein and with a measured radius of gyration. Inherent in this search is the assumption that the protein has a globular shape that can be well approximated by an ellipsoid. If this assumption is correct, then a unique best-fit solution can be determined within the constrained model set. Protein complexes can be similarly modeled using two-ellipsoid models. In general, as you add more variables to the model (e.g., more independent shapes), you will need more data against which to test the model if you are to find a unique solution even within a constrained set. The more contrast points you have in a neutron contrast variation data set, the better you are able to test more complicated models. There are no agreed upon consensus guidelines for modeling small-angle scattering data, and therefore it is reader beware in evaluating models in terms of the underlying assumptions and constraints, as well as the number of independent data sets being used in the testing. The TnC–TnI models discussed in **Subheading 3.5.2.** provide a good example of where difficulties can arise. These proteins are not "typical" globular proteins and they appear be intertwined in a complicated manner. In contrast, there is ample evidence that the CaM/MLCK complexes can be well approximated using elliptical shapes. As a result, these proteins have proven much more straightforward to interpret.

## Acknowledgments

The work described in this chapter was supported by the Department of Energy's Office of Biological and Environmental Research (DOE/OBER) and

by a National Institutes of Health (NIH) grant (GM40528). The scattering experiments described used X-ray facilities at the Stanford Synchrotron Radiation Laboratory supported by DOE/OBER and an NIH Research Resource Grant, and neutron facilities at the Department of Commerce National Institute of Standards Technology supported by a National Science Foundation Grant.

## References

1. Glatter, O. and Kratky, O. (1982) *Small Angle X-ray Scattering*. Academic, London and New York.
2. Feigin, L. A. and Svergun, D. I. (1987) *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. Plenum, New York and London.
3. Trewhella, J., Gallagher, S. C., Krueger, J. K., and Zhao, J. (1998) Neutron and x-ray solution scattering provide insights into biomolecular structure and function. *Science Progress* **81,** 101–122.
4. Wu, C.-F. and Chen, S.-H. (1988) Small angle neutron and x-ray scattering studies of concentrated protein solutions II. Cytochrome c. *Biopolymers* **27,** 1065–1083.
5. Krigbaum, W. R. and Kugler, F. R. (1970) Molecular conformation of egg-white lysozyme and bovine (-lactalbumin in solution. *Biochemistry* **9,** 1216–1223.
6. Guinier, A. and Fournet, G. (1955) *Small-Angle Scattering of X-rays*. Wiley, New York.
7. Moore, P. B. (1980) Small angle scattering information content and error analysis. *J. Appl. Cryst.* **13,** 168–175.
8. Glatter, O. (1979) The interpretation of real-space information from small-angle scattering experiments. *J. Appl. Cryst.* **12,** 166–175.
9. Svergun, D. I., Semenyuk, A. V., and Feigen, L. A. (1988) Small-angle scattering data treatment by the regularization method. *Acta Cryst.* **A44,** 244–250.
10. Heidorn, D. B. and Trewhella, J. (1988) Comparison of the crystal and solution structures of calmodulin and troponin C. *Biochemistry* **27,** 909–915.
11. Krueger, J. K., Olah, G. A., Rokop, S. E., Zhi, G., Stull, J. T., and Trewhella, J. (1997) The structure of $4Ca^{2+}$·calmodulin and a functional myosin light chain kinase in the activated complex. *Biochemistry* **36,** 6017–6023.
12. Krueger, J., Zhi, G., Stull, J. T., and Trewhella, J. (1998) Neutron scattering studies reveal further details of the $Ca^{2+}$/calmodulin-dependent activation mechanism of myosin light chain kinase. *Biochemistry* **37,** 13,997–14,004.
13. Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., and Sowadski J. M. (1991) Crystal structure of the catalytic subunit of cAMP-dependent protein kinase. *Science* **253,** 407–414.
14. Krueger, J. K., Padre, R. C., and Stull, J. T. (1995) Intrasteric regulation of myosin light chain kinase. *J. Biol. Chem.* **270,** 16,848–16,853.
15. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R., and Cook, W. J., (1985) Three-dimensional structure of calmodulin. *Nature* **315,** 37–40.
16. Ikura, M. (1996) Calcium binding and conformational response ion EF-hand proteins. *Trends Biochem. Sci.* **21,** 14–17.

17. Seaton, B. A., Head, J. F., Engelman, D. M., and Richards, F. M. (1985) Calcium-induced increase in the radius of gyration and maximum dimension of calmodulin measured by small-angle scattering. *Biochemistry* **24,** 6740–6743.
18. Trewhella, J. (1992) The solution structures of calmodulin and its complexes with peptides based on target enayme binding domains. *Cell Calcium* **13,** 407–420.
19. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256,** 632–638.
20. Persechini, A., Gansz, K. J., and Paresi, R. J. (1996) A role in enzyme activation for the N-terminal leader sequence in calmodulin. *J. Biol. Chem.* **271,** 19,279–19,282.
21. Krueger, J., Bishop, N. A., Blumenthal, D. K., Zhi, G., Beckingham, K., Stull, J. T., and Trewhella, J. (1998) Calmodulin binding to myosin light chain kinase begins at substoichiometric $Ca^{2+}$ concentrations: a small-angle scattering study of binding and of conformational transitions. *Biochemistry* **37,** 17,810–17,817.
22. Bayley, P., Findlay, W. A., and Martin, S. R. (1996) Target recognition by calmodulin: dissecting the kinetics and affinity of interaction using short peptide sequences. *Protein Sci.* **5,** 1215–1228.
23. Peerson, O. B., Madson, T. S., and Falke, J. J. (1997) Intermolecular tuning of calmodulin by target peptides and proteins: differential effects on $Ca^{2+}$ binding and implications for kinase activation. *Protein Sci.* **6,** 794–807.
24. Trewhella, J., Blumenthal, D. K., Rokop, S. E., and Seeger, P. A. (1990) Small-angle scattering studies show distinct conformations of calmodulin in its complexes with two peptides based on the regulatory domain of the catalytic subunit of phosphorylase kinase. *Biochemistry* **29,** 9316–9324.
25. Olah, G. A., Rokop, S. E., Wang, C.-L. A., Blechner, and S. L., Trewhella, J. (1994) Troponin I encompasses an extended troponin C in the $Ca^{2+}$ bound complex: a small-angle x-ray and neutron scattering study. *Biochemistry* **33,** 8233–8239.
26. Stone, D. B., Timmins, P. A., Schneider, D. K., Krylova, I., Ramos, C. H. I., Reinach, F. C., and Mendelson, R. A. (1998) The effect of regulatory $Ca^{2+}$ on the in situ structures of troponin C and troponin I: a neutron scattering study. *J. Mol. Biol.* **281,** 689–704.
27. Olah, G. A. and Trewhella, J. (1994) A model structure of the muscle protein complex $4Ca^{2+}$·troponin C·troponin I derived from small-angle scattering data: implications for regulation. *Biochemistry* **33,** 12,800–12,806.
28 Tung, C.-S., Wall, M. E., Gallagher, S. C., and Frewella, J. (2000) A model of troponin-I in complex with troponin-C using hybrid experimental data: The inhibitory region is a B-hairpin. *Protein Science* **9,** 1312–1326.
29. Zaccai, G. and Jacrot, B. (1983) Small angle neutron scattering. *Ann Rev. Biophys. Bioeng.* **12,** 139–157.

# 13

# Investigation of Calcium-Binding Proteins Using Electrospray Ionization Mass Spectrometry

**Amanda L. Doherty-Kirby and Gilles A. Lajoie**

## 1. Introduction

Electrospray ionization mass spectrometry (ESI-MS) is a soft ionization technique that is rapid and more sensitive than many other available techniques that are used to characterize macromolecules. It is particularly suitable for studying proteins in their native state as the solvent conditions for transferring and ionizing protein molecules from solution to the gas phase can be similar to those used for characterizing proteins in solution. Following a study of myoglobin at the beginning of decade ***(1)***, there have been a number of reported ESI-MS studies of specific noncovalent protein-ligand complexes. A recent review by Loo provides a thorough list of ESI-MS studies of proteins interacting with metals, small molecules, peptides and proteins, and nucleic acids ***(2)***.

ESI-MS has a number of advantages when compared to other available physical methods ***(2)***. Both NMR and X-ray crystallography require milligrams of material and are fairly slow techniques, whereas ESI-MS is rapid and can detect picomole to femtomole quantities of protein. The high concentrations of protein required for NMR may lead to precipitation, which is less likely to occur with the much lower concentrations used for ESI-MS.

Other physical methods, such as circular dichroism, fluorescence, and UV absorption, are indirect as they monitor overall conformational changes in the protein of interest upon binding to their ligands. These studies can be ambiguous because of such factors as incomplete demetallation and failure to determine accurate molar ratios ***(3)***. One major advantage of ESI-MS is its ability to resolve relative amounts of coexisting species. Therefore, ESI-MS is completely amenable to titration studies, which can provide information about

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

metal ion partitioning and the sequence of metal binding. To date, the determination of stoichiometry of calcium binding is the most documented application of ESI-MS in the study of metal-binding proteins.

Standard "denaturing" ESI-MS conditions used for determination of protein molecular weight are solutions of pH 2.0 to 4.0 in the positive ion mode and 8.0 to 10.0 in the negative ion mode to yield the best sensitivity. An organic cosolvent, such as acetonitrile or methanol, is often used to enhance sensitivity and signal stability. To study protein interactions by ESI-MS, nondenaturing conditions including a volatile buffer, such as ammonium acetate or ammonium carbonate close to neutral pH, and low temperature are used. These milder conditions lead to a significant decrease in sensitivity. Mass spectrometer parameters, such as capillary temperature, ion mode, and voltage, must be optimized for each system. This requires a balance between maintaining the intact complex while adjusting parameters for sufficient ion desolvation and ionization.

A set of criteria should be met to provide evidence that the interactions observed are specific ***(4)***. First, for tight binding complexes, the predominant species should correspond to that identified for the protein in solution. The intensity of signal corresponding to complexes should be altered by changes in instrumental conditions, such as increased capillary temperature or applied voltage. Mass spectra should reflect differences in complex formation and be sensitive to solution conditions, such as pH, temperature, type and concentration of buffer components. Structural modification of the complex components either in the protein or the ligand should alter the relative intensity of observed species corresponding to increased or decreased binding in solution. ***(4)***.

This chapter is a short review of studies on noncovalent interactions of calcium-binding proteins that have been characterized by ESI-MS, followed by some detail of our own work on calmodulin (CaM). The studies of calcium-binding proteins in their native state using ESI-MS can be grouped into three types:

1. Properties of calcium binding (i.e., stoichiometry and cooperativity) determined through examination of the metal-bound species present in mass spectra and comparison to spectra of the metal free protein.
2. Conformational changes detected by a shift in the mass-to-charge (*m/z*) envelope or with the use of hydrogen–deuterium exchange.
3. Interactions of calcium-binding proteins with other molecules.

CaM is, by far, the most extensively studied $Ca^{2+}$-binding protein. A number of groups have determined by ESI-MS that CaM binds four calcium ions, consistent with the observed stoichiometry in solution ***(3,5–8)***. Studies with calcium, magnesium, and terbium also showed that there were additional low-affinity binding sites ***(7)***.

As stated earlier, the nature of the measurements obtained with ESI-MS yields precise information about stoichiometry. This is illustrated by an ESI-MS study of the EF-hand protein calbindin $D_{28K}$, which has been shown using to bind between three and six calcium ions by various solution methods. ESI mass spectra showed clearly that four calcium ions are bound to calbindin $D_{28K}$ as a mass difference of 151 Da (expected = 4 × 40 Da (mass of $Ca^{2+}$) – 4 × 2 Da (loss of protons) = 152 Da) was observed between the apo and the holo forms of the protein. Mutants that contained deletions of one or two of the EF hands were used to help localize the sites of metal binding ***(9)***.

A recent study of human parvalbumin provides a good example of sensitivity to structural modifications. Variants that had mutations in both of the two EF hands did not bind any $Ca^{2+}$ acting as a suitable negative control and indicating that the interactions observed between the parvalbumins and calcium were indeed specific. Wild-type protein bound two $Ca^{2+}$, whereas the proteins that had mutations in one EF hand bound one calcium ***(10)***. This method was also used to study parvalbumin extracted from murine and rat tissue samples ***(10)***. Additional calcium-binding proteins for which $Ca^{2+}$ stoichiometry has been determined by ESI-MS include parvalbumin ***(3,5,11)***, lactalbumin ***(5)***, matrylisin ***(12)***, and the stromelysin catalytic domain ***(5)***.

ESI-MS has been used to study cooperativity of binding and to compare binding affinity for different metals ions. Typically, a metal-free protein is titrated with a metal ion and the ESI mass spectra are obtained at various metal concentrations. Examination of the relative amounts of co-existing species at each concentration yields information about preferential binding. For example, if a protein has two $Ca^{2+}$-binding sites that are both high affinity and independent, the mass spectrum should reveal the presence of the mixture apo-protein: $Ca^{2+}$-protein: $(Ca^{2+})_2$-protein in a 1:2:1 ratio when one equivalent of $Ca^{2+}$ (with respect to the protein) is present. If the sites were independent, but one had a higher affinity, the mass spectrum should primarily show a species with only one $Ca^{2+}$ bound. Alternatively, if the binding was positively cooperative, one would expect that the spectrum would show some apo-protein and $(Ca^{2+})_2$-protein with little $(Ca^{2+})_1$-protein present. This is illustrated in a study of the binding of $Ca^{2+}$ and $Cd^{2+}$ to wild-type and mutant calbindin $D_{9K}$ ***(13)***. The spectrum of the wild-type protein indicates that as the calcium concentration is increased, the calbindin proceeds from being in the apo form to binding two $Ca^{2+}$ with the $(Ca^{2+})_1$-species present in very small amounts. The mutant N56A calbindin, when titrated with calcium acetate, shows high amounts of apo-calbindin, $(Ca^{2+})_1$-calbindin and $(Ca^{2+})_2$-calbindin depending on the $Ca^{2+}$ concentration. This type of spectrum was also observed for the binding of $Cd^{2+}$ to both the wild-type and mutant proteins. Thus, wild-type calbindin $D_{9K}$ binds

two $Ca^{2+}$ cooperatively, the mutant binds $Ca^{2+}$ sequentially (i.e., one high-affinity and one low-affinity site), and both proteins bind $Cd^{2+}$ sequentially.

Conformational changes upon metal binding can be detected in two different ways using ESI-MS. The first is the observation of a shift of the *m/z* envelope to lower charge states (i.e., higher *m/z*). This phenomenon has been observed in several studies of calcium-binding proteins and speculated to be a result of alterations in conformation ***(5,7,12,14)***. This was confirmed in a recent study on calbindin $D_{28K}$. ***(15)***. The study with calbindin $D_{28K}$ showed there is a correlation of the change in the *m/z* envelope to the changes observed in fluorescence and near UV-CD spectra with titration of $Ca^{2+}$, although far UV-CD studies indicated little change in secondary structure. Thus, the shift to higher *m/z* (i.e., less charge) is indicative of alterations in the tertiary structure of the protein rather than in secondary structure. The calbindin $D_{28K}$ study is an excellent example of using ESI-MS as a very convenient screening method to determine if the binding of calcium (or other metals) to proteins lead to conformational changes.

The second approach to examine conformational change is the use of hydrogen–deuterium exchange with ESI-MS. The basic experiment involves incubating the protein in volatile buffer, which is prepared with deuterium oxide ($D_2O$) with or without metal present, lowering the pH to 3.0 with the addition of acid, and digesting the protein with an acid protease (typically pepsin). The proteolysis products are then analyzed by ESI-MS, usually with a separation step prior to mass spectrometry. Conformational changes are determined as an increase or a decrease in the rate of deuterium exchange for a particular region of the protein as observed by the increase or decrease in the mass of a particular peptide fragment in the ESI-MS spectrum. Only one calcium-binding protein (recoverin) has been studied in this manner ***(16)***. The technique has also been applied to protein-ligand binding for several proteins ***(17–20)***; to compare amide exchange rates in FK506-binding protein determined by both ESI-MS and NMR ***(21)***, and to compare native and nonnative states of cytochrome c ***(20)***.

Recoverin, an *N*-terminal myristoylated protein, involved in visual signal transduction is thought to contain a "calcium-myristoyl switch" where $Ca^{2+}$-binding results in extension of the acyl group allowing it to interact with membranes and/or other proteins. Three peptides corresponding to regions of the hydrophobic myristoyl-binding pocket of acylated recoverin show destabilization (i.e., increased hydrogen–deuterium exchange rates) in the presence of $Ca^{2+}$. One of these peptides was stabilized by $Ca^{2+}$ when nonmyristoylated recoverin was used. This evidence suggests that when $Ca^{2+}$ is not bound, the acyl group protrudes into the hydrophobic binding pocket and excludes solvent. In the absence of calcium, the nonmyristoylated recoverin may transiently unfold and have a less stable structure than myristoylated recoverin ***(16)***. This technique, although more time consuming than observation of a shift in the

*m/z* envelope, can be used to locate regions that are most affected by a conformational change upon metal binding.

Interaction of calcium-binding proteins with other molecules is ubiquitous in nature. Upon binding $Ca^{2+}$, CaM binds and activates numerous enzymes. Not surprisingly, CaM has been the most widely studied calcium-binding protein by ESI-MS. The interaction of calmodulin with both small pyrazine derivatives and with the peptides melittin and CaM kinase II have been examined by ESI-MS ***(23–25)***. The binding of the two peptides by CaM was shown to be calcium-dependent and in a 1:1 stoichiometric fashion. Veenstra et al. have shown the interaction to be source temperature dependent supporting that the interaction is truly specific ***(25)***. Homodimers of S100B and of CaM have also been observed in the gas phase ***(26,27)***.

One non-EF-hand $Ca^{2+}$-binding protein that has been studied using ESI-MS is matrilysin, a matrix metalloprotease. It has been determined by ESI-MS that the minimum requirement of matrilysin to bind small molecule inhibitors is to also have two $Zn^{2+}$ and one $Ca^{2+}$ bound. In optimal conditions, matrilysin binds two of each metal. Additionally, Feng et al. found that the relative ion intensities of complexes formed between matrilysin and two inhibitors correlated with their solution binding constants ***(12)***.

Our own work has focused on the interaction of CaM with a variety of peptides corresponding to binding regions of target proteins. The peptides studied range from those that show strong binding to those that interact much more weakly. We have also examined the interaction of the isolated lobes of CaM, $TR_1C$ (CaM 1-75), and $TR_2C$ (CaM 78-148) with several of these stronger binding peptides.

## 2. Materials

1. Recombinant bovine CaM, $TR_1C$, $TR_2C$, and a peptide corresponding to the calmodulin-binding region of skeletal myosin light chain kinase (MLCK) were obtained from H. Vogel (University of Calgary). The sequence of the MLCK peptide is KRRWKKNFIAVSAANRFKKISS.
2. 4.5 m*M* ammonium acetate [$NH_4$(OAc)] prepared using HPLC grade $H_2O$ (Milli-Q) and adjusted to pH 6.8 using acetic acid. Prepare fresh or store refrigerated in a Nalgene bottle (*see* **Note 1**).
3. HPLC-grade methanol (MeOH).
4. Infusion pump (Type 365, Sage Instruments) with gas-tight Hamilton syringe to give low (5 μL/min) flow of sample. Alternatively, an HPLC pump that can deliver a low, pulse-free flow of solvent can be used.
5. Calmodulin, $TR_1C$, and $TR_2C$ stock solutions (0.1 m*M*) in 4.5 m*M* $NH_4$(OAc) prepared fresh or stored at 4°C and brought to room temperature about 1 h before use.
6. Calcium acetate stock solution (1 m*M*) prepared fresh before use.

7. Peptide stock solutions in 4.5 m*M* $NH_4(OAc)$ prepared fresh or stored at 4°C and brought to room temperature before use.
8. Calibrant — myoglobin (Sigma) (*see* **Note 2**). Stock solutions should be stored frozen and thawed prior to use.

## 3. Methods

### 3.1. Mass Spectrometry

The analyses were performed on samples containing calmodulin at a concentration of 10 picomoles/microliter (10 μ*M*) and calcium acetate a concentration of 200 μ*M* in 90:10 4.5 m*M* $NH_4(OAc)$-MeOH.

#### 3.1.1. Sample Preparation

1. Add CaM such that final working sample concentration will be 10 μ*M*.
2. Add suitable amount of calcium acetate stock solution such that the final concentration will be 200 μ*M*.
3. Add peptide to the desired concentration. Typically this was 1.0 equivalent with respect to CaM or CaM fragment. Initial concentrations of peptides and proteins were determined by UV analysis.
4. Dilute sample with buffer to the desired final concentrations (*see* **Note 3**).
5. Incubate at room temperature for 30 min.
6. Load into gas-tight syringe for infusion (*see* **Note 4**).

#### 3.1.2. Mass Spectrometer Parameters (see ***Note 5***)

Spectra were obtained in negative ion mode as this had less interference with nonspecific adducts than positive ion mode. Negative ion mass spectra were acquired on a Micromass Quattro II triple quadrupole mass spectrometer (Micromass, UK with a *m/z* up to 4 kDa.

1. Set the cone voltage to –55 V, capillary voltage to –2.8 kV, HV lens to –0.5 kV, and the capillary temperature to 55°C (*see* **Note 6**).
2. Acquire data by scanning over the *m/z* range from 700 to 3000 in 8 s (*see* **Note 7**).
3. Using the MCA (Multi-Channel Analyzer) mode of data collection, sum the data over 10–15 min.

#### 3.1.3. Processing the Data

Deconvoluted spectra were obtained using the Maximum Entropy ***(28)*** software supplied with the manufacturer's software (MassLynx v. 2.0).

1. Subtract of baseline from raw data.
2. Obtain a survey spectrum using MaxEnt. The width at half height parameter should be set for a singlet peak around the center of the *m/z* distribution (*see* **Note 8**). The output range is generally set to 15–30 kDa with a resolution of 10 Da.
3. Obtain a more accurate deconvoluted spectrum using a range to include the components observed in the survey spectrum and a resolution of 1 Da. Let the MaxEnt analysis proceed to convergence (*see* **Note 9**).

4. Centroid the data and obtain a centered spectrum with the heights corresponding to the areas under the peaks to obtain relative intensity data (*see* **Note 10**).

## 3.2. Results

### 3.2.1. Free Calmodulin

**Figure 1A** shows the *m/z* distribution for apo-CaM obtained under nondenaturing conditions, whereas **Fig. 1B** shows a similar spectrum for $Ca^{2+}$–CaM in the presence of 200 μ*M* $Ca^{2+}$ using the same conditions. Unlike the apo-CaM, two *m/z* distributions are observed for $Ca^{2+}$-bound CaM. This is thought to correspond to two different forms of CaM being present, one with a more native structure (major) and one with an extended denatured-like form (minor). The MaxEnt spectra (*see* **Fig. 1C**,**D)** indicates a mass difference of 152 Da between the major species with $Ca^{2+}$ present and apo-Cam corresponding to four $Ca^{2+}$ bound (4 × 40 Da – 4 × 2Da = 152 Da) as expected.

### 3.2.2. Observation of Peptide Complexes

The original *m/z* data and the MaxEnt deconvoluted spectra for the 1:1 MLCK peptide/$(Ca^{2+})_4$–CaM are shown in **Fig. 2A**,**B**, respectively. The *m/z* distribution shows three major species corresponding to the –7, –8, and –9 charge states of the complex. The observed mass of 19485 Da indicates a 1:1 complex between $(Ca^{2+})_4$–CaM and the MLCK peptide is formed. There is also no free CaM apparent in the spectrum indicating that CaM binds the MLCK peptide very tightly. This is consistent with a solution dissociation constant of 2 n*M* that has been reported for his complex ***(29)***.

### 3.2.3. Interaction of MLCK with Isolated Lobes of CaM

In order to determine which lobe of CaM interacts more strongly with MLCK, ESI-MS was performed on 1:1 mixtures of the individual lobes with the MLCK peptide. Deconvoluted spectra are shown in **Figs. 3A**,**B**. In each case, complexes with a stoichiometry of 1:1 $(Ca^{2+})_2$–CaM fragment:MLCK peptide is observed. Comparison of the data for $TR_1C$ + MLCK (*see* **Fig. 3A**) and $TR_2C$ + MLCK (*see* **Fig. 3A**) shows clearly that the MLCK peptide binds more strongly to $TR_2C$ than to $TR_1C$. In a separate experiment, we confirmed that both the free lobes $TR_1C$ and $TR_2C$ ionize similarly. The intensity of the complexed vs free $TR_1C$ or $TR_2C$ corresponds to a $K_d$ of 25 μ*M* for $TR_1C$ and of 0.58 μ*M* for $TR_2C$. Neither of these complexes is observed using gel band-shift assays consistent with them having dissociation constants greater than 0.2 μ*M*. These results also correlate with the relative strength of the binding of the calmodulin fragments to rabbit skeletal MLCK ($K_d$ = 0.3 μ*M* for $TR_2C$–MLCK and 3 μ*M* for $TR_1C$–MLCK) ***(30)***. Alteration of the instrumental parameters (temperature and cone voltage) for $TR_2C$ + MLCK produced supporting data

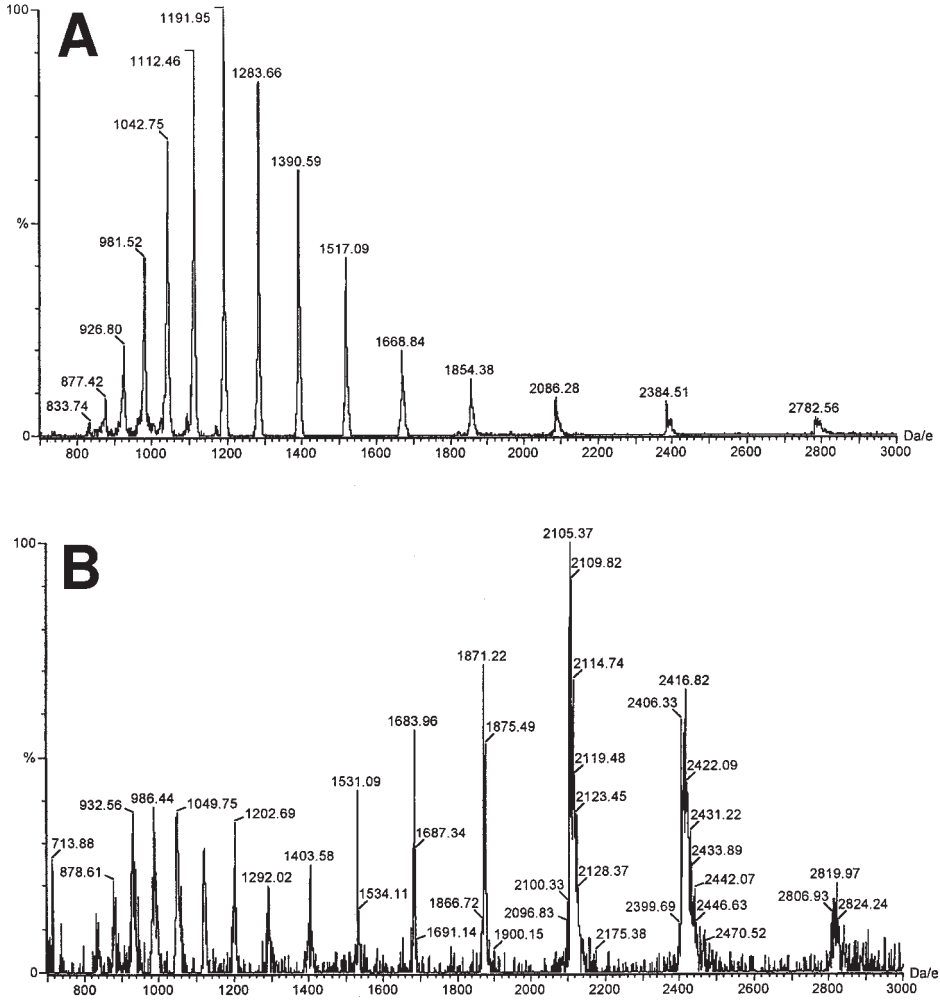


Fig. 1. Baseline subtracted and smoothed data for **(A)** apo-CaM and **(B)** CaM with 200 μ*M* calcium acetate present.

for the specific nature of this complex. Similar ESI-MS experiments have been performed on a range of CaM-binding peptides. The intensity of the complex vs free CaM reflects closely the $K_d$ measured in solution (full study to be published elsewhere).

### 3.3. Conclusions

ESI-MS is an effective tool for studying the properties of calcium-binding proteins. Information about stoichiometry, cooperativity, conformation, and

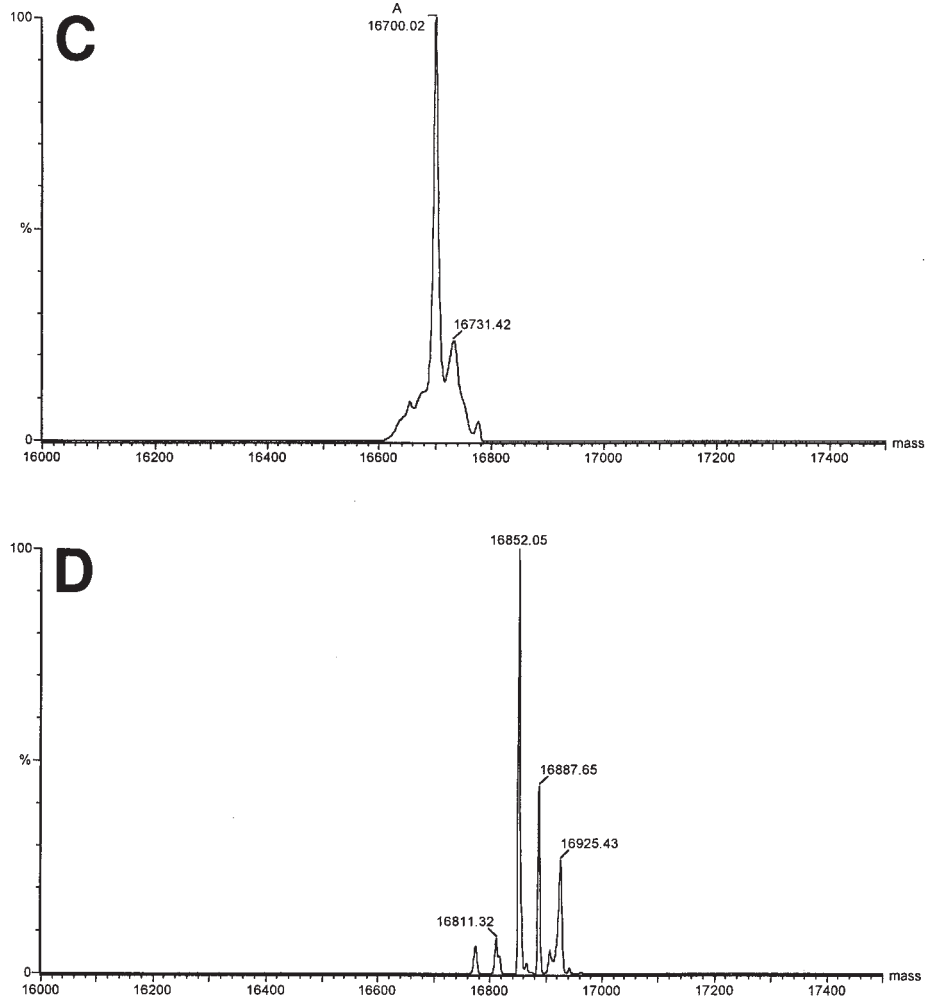


Fig. 1. (Deconvoluted spectra for **(C)** apo-CaM and **(D)** $Ca^{2+}$-loaded CaM). The expected molecular weight for apo-CaM is 16,700 Da, while that for $(Ca^{2+})_4$-CaM is 16,852 Da.

interactions with other molecules is readily attainable with small amounts of material. From the data presented here, we have shown that the tight-binding interaction between CaM and MLCK is readily observed in the gas phase and that $Tr_2C$ binds MLCK more tightly than $TR_1C$. From our own studies, it is known that the observed relative ion intensities correlate well with solution binding constants for the system of CaM and CaM fragments interacting with target peptides. Thus, the detailed information obtained by ESI-MS studies

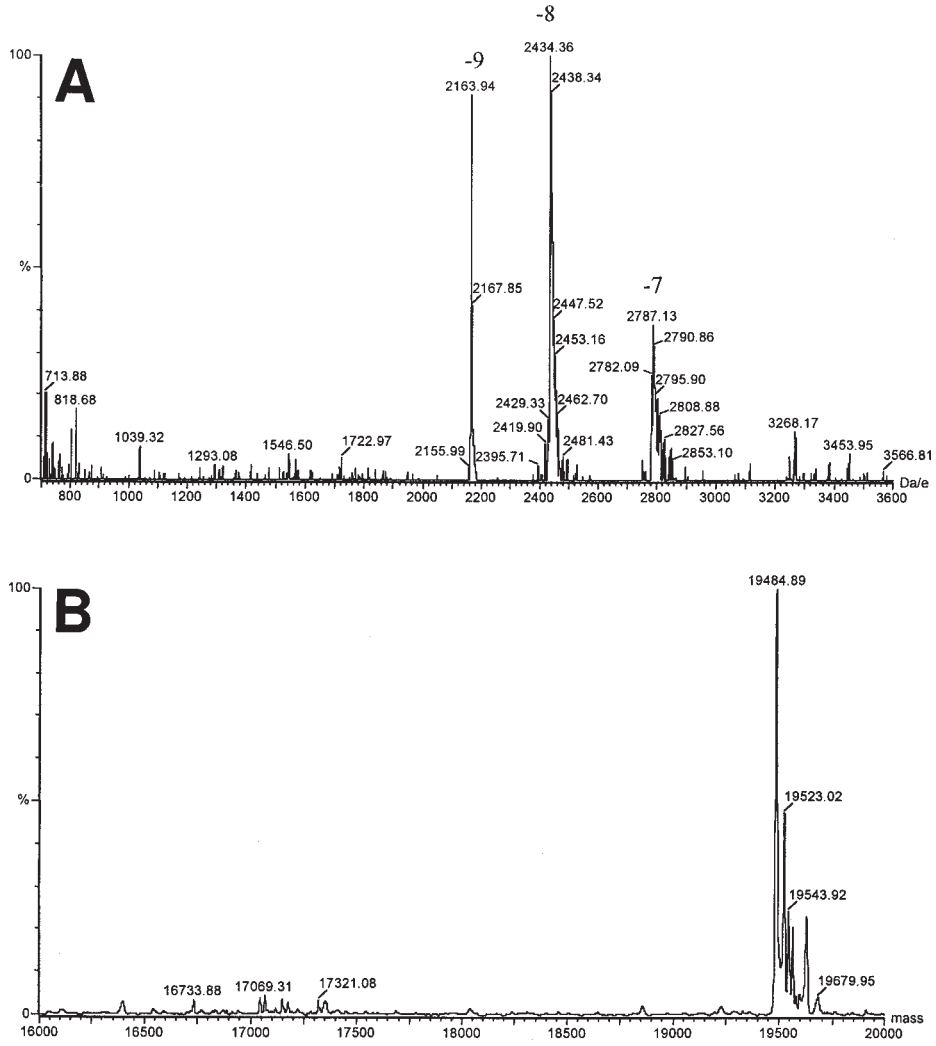


Fig. 2. **(A)** Original *m/z* data (baseline subtracted and smoothed) and **(B)** deconvoluted spectrum for the $(Ca^{2+})_4$-CaM:MLCK peptide complex. Expected molecular weight for the 1:1 complex is 19,486 Da.

ensures that mass spectrometry will play a significant role in the elucidation of the properties of metal-binding proteins.

## 4. Notes

1. To remove $Ca^{2+}$ from water prior to use, run through Chelex-100. Store in Nalgene bottles that have been cleaned with 10% nitric acid.

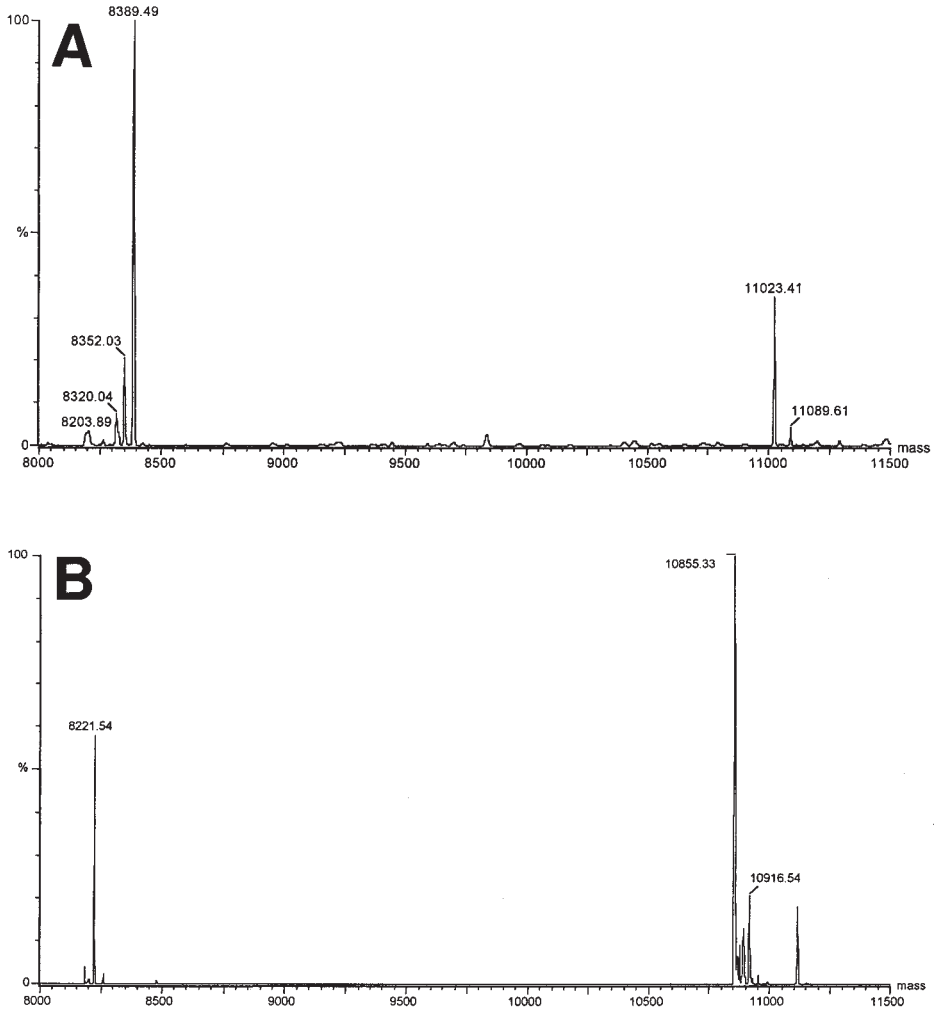


Fig. 3. Deconvoluted spectra of 1:1 mixtures of **(A)** $(Ca^{2+})_2$-$TR_1C$ and MLCK peptide and **(B)** $(Ca^{2+})_2$-$TR_2C$ and MLCK peptide. Expected masses are $(Ca^{2+})_2$-$TR_1C$, 8390 Da, $(Ca^{2+})_2$-$TR_1C$:MLCK peptide, 11,024 Da, $(Ca^{2+})_2$-$TR_2C$, 8221 Da, $(Ca^{2+})_2$-$TR_2C$:MLCK peptide, 10,855 Da.

2. Calibration is performed via a separate injection of the calibrant. Calibration was performed using denaturing conditions (50:50 acetonitrile-water, 1% $NH_4OH$) in negative ion mode.
3. MeOH to 10% of the total volume was added after the 30 min incubation period.
4. If complexes exhibit precipitation, centrifuge or filter before loading into syringe.

5. Although some of the terms apply to all instruments with an electrospray source, some terms apply only to Micromass instruments. For example the cone voltage refers to the sampling orifice to skimmer potential difference on our instrument.
6. These parameters will vary depending on the system being studied, but worked well for CaM binding to target peptides. Parameters (the most important are capillary temperature and cone voltage) are set to allow efficient desolvation and ionization without the destruction of the complex.
7. This range encompasses most of the species observed in our study. If, in an initial experiment, it appeared that another charge state might be evident at higher *m/z*, the scan range was increased.
8. Process only the part of the spectra that contain multiply charged data. Refer to the manufacturer's instruction manual for a full determination of the width at half height parameter that is required for MaxEnt.
9. MaxEnt data must converge in order to obtain quantitative data. The area under the peak correlates to the relative amounts of species present. Additional species corresponding to adducts of the proteins may be observed. For this study, we mainly observed addition of +17 ($NH_4^+$) and +38 ($Ca^{2+}$) with +22 ($Na^+$) being observed occasionally.
10. Centering must be done based on area in order to obtain quantitative data about the relative abundance of species present in the spectra. This is based on the assumption that the species ionize similarly. Calculations of $K_d$ were based on the relative heights of the centered spectra for the complex being studied using

$$K_d = [\text{Free CaM (or fragment)}][\text{Free Peptide}]/[\text{complex}].$$

## References

1. Katta, V. and Chait, B. T. (1991) Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* **113,** 8534–8535.
2. Loo, J. A. (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16,** 1–23.
3. Hu, P. and Loo, J. A. (1995) Determining calcium-binding stoichiometry and cooperativity of parvalbumin and calmodulin by mass spectrometry. *J. Mass Spectrom.* **30,** 1076–1082.
4. Smith, R. D. and Light-Wahl, K. J. (1993) The observation of non-covalent interactions in solution by electrospray ionization mass spectrometry: promise, pitfalls and prognosis. *Biol. Mass Spectrom.* **22,** 493–501.
5. Hu, P., Ye, Q. Z., and Loo, J. A. (1994) Calcium stoichiometry determination for calcium binding proteins by electrospray ionization mass spectrometry. *Anal. Chem.* **66,** 4190–4194.
6. Lafitte, D., Capony, J. P., Grassy, G., Haiech, J., and Calas, B. (1995) Analysis of the ion binding sites of calmodulin by electrospray ionization mass spectrometry. *Biochemistry* **34,** 13,825–13,832.

7. Lafitte, D., Capony, J. P., Grassy, G., Haiech, J., and Calas, B. (1995) Electrospray ionization mass spectrometric study of calcium binding to calmodulin in the presence of magnesium and terbium. *J. Mass Spectrom. Rapid Commun. Mass Spectrom.* S192–S196.
8. Johnson, K. L., Veenstra, T. D., and Tomlinson, A. J. (1997) Determination of noncovalent metal ion/protein interactions using a microflow electrospray ionization mass spectrometry interface. *Rapid Commun. Mass Spectrom.* **11,** 939–942.
9. Veenstra, T. D., Johnson, K. L., Tomlinson, A. J., Naylor, S., and Kumar, R. (1997) Determination of calcium-binding sites in rat brain calbindin $D_{28K}$ by electrospray ionization mass spectrometry. *Biochemistry* **36,** 3535–3542.
10. Troxler, H., Kuster, T., Rhyner, J. A., Gehrig, P., and Heizmann, C. W. (1999) Electrospray ionization mass spectrometry:analysis of the $Ca^{2+}$-binding properties of human recombinant α-parvalbumin and nine mutant proteins. *Anal. Biochem.* **268,** 64–71.
11. Hu, P., Buckel, S. D., Whitton, M. M., and Loo, J. A. Calcium binding properties and sequence of frog (*Rana tegrinka*) parvalbumin as determined by electrospray ionization-mass spectrometry. (1996) *Eur. Mass Spectrom.* **2,** 69–76.
12. Feng, R., Castelhano, A. L., Billedeau, R., and Yuan, Z. (1995) Study of non-covalent enzyme-inhibitor complexes and metal binding stoichiometry of matrilysisn by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **6,** 1105–1111.
13. Chazin, W. and Veenstra, T. D. (1999) Determination of the metal-binding cooperativity of wild-type and mutant calbindin $D_{9K}$ by electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **13,** 548–555.
14. Veenstra, T. D., Johnson, K. L., Tomlinson, A. J., Naylor, S., and Kumar, R. (1997) Electrospray ionization mass spectrometry temperature effects on metal ion: protein stoichiometries and metal induced conformational changes in calmodulin. *Eur. Mass Spectrom.* **3,** 453–459.
15. Veenstra, T. D., Johnson, K. L., Tomlinson, A. J., Kumar, R., and Naylor, S. (1998) Correlation of fluorescence and circular dichroism spectroscopy with electrospray ionization mass spectrometry in the determination of tertiary conformational changes in calcium-binding proteins. *Rapid Commun. Mass Spectrom.* **12,** 613–619.
16. Neubert, T. A., Walsh, K. A., Hurley, J. B., and Johnson, R. S. (1997) Monitoring calcium-induced conformational changes in recoverin by electrospray mass spectrometry. *Protein Sci.* **6,** 843–850.
17. Johnson, R. S. and Walsh, K. A. (1994) Mass spectrometric measurement of protein amide hydrogen exchange rates of apo- and holo-myoglobin. *Protein Sci.* **3,** 2411–2418.
18. Wang, F., Blanchard, J. S., and Tang, X. J. (1997) Hydrogen exchange/electrospray ionization mass spectrometry studies of substrate and inhibitor binding and conformational changes of *Escherichia coli* dihydrodipicolinate reductase. *Biochemistry* **36,** 3755–3759.
19. Wang, F., Li, W., Emmett, M. R., Hendrickson, C. L., Marshall, A. G., Zhang, Y. L., et al. (1998) Conformational and dynamic changes of Yersinia protein tyrosine

phosphatase induced by ligand binding and active site mutation and revealed by H/D exchange and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Biochemistry* **37,** 15,289–15,299.

20. Wang, F., Scapin, G., Blanchard, J. S., and Angeletti, R. H. (1998) Substrate binding and conformational changes of *Clostridium glutamicum* diaminopimelate dehydrogenase revealed by hydrogen/deuterium exchange and electrospray mass spectrometry. *Protein Sci.* **7,** 293–299.
21. Zhang, Z., Li, W., Logan. T. M., Li, M., and Marshall, A. G. (1997) Human recombinant [C22A]FK506-binding protein amide hydrogen exchange rates from mass spectrometry match and extend those from NMR. *Protein Sci.* **6,** 2203–2217.
22. Maier, C., Kim, O. H., and Deinzer, M. L. (1997) Conformational properties of the A-state of cytochrome c studied by hydrogen/deuterium exchange and electrospray mass spectrometry. *Anal. Biochem.* **252,** 127–135.
23. Lafitte, D., Benezech, V., Bompart, J., Laurent, F., Bonnet, P. A., Chapat, J. P., et al. (1997) Characterization of low affinity complexes between calmodulin and pyrazine derivatives by electrospray ionization mass spectrometry. *J. Mass Spectrom.* **32,** 87–93.
24. Nemirovskiy, O. V., Ramanathan, R., and Gross, M. L. (1997) Investigation of calcium-induced, noncovalent association of calmodulin with melittin by electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **8,** 809–812.
25. Veenstra, T. D., Tomlinson, A. J., Benson, L., Kumar, R., and Naylor, S. (1998) Low temperature aqueous electrospray ionization mass spectrometry of noncovalent complexes. *J. Am. Soc. Mass Spectrom.* **9,** 580–584.
26. Raferty, M. J. and Geczy, C. L. (1998) Identification of noncovalent dimeric complexes of the recombinant murine S100 protein CP10 by electrospray ionization mass spectrometry and chemical cross-linking. *J. Am. Soc. Mass Spectrom.* **9,** 533–539.
27. Lafitte, D., Heck, A. J. R., Hill, T. J., Jumel, K., Harding, S. E., and Derrick, P. J. (1999) Evidence of noncovalent dimerization of calmodulin. *Eur. J. Biochem.* **261,** 337–344.
28. Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A., and Ski lling, J. (1992) Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* **6,** 707–711.
29. Ohki, S., Ikura, M., and Zhang, M. (1997) Identification of $Mg^{2+}$-binding sites and the role of $Mg^{2+}$ on target recognition by calmodulin. *Biochemistry* **36,** 4309–4316.
30. Persechini, A., McMillan, K., and Leakey, P. (1994) Activation of myosin light chain kinase and nitric oxide synthase activities by calmodulin fragments. *J. Biol. Chem.* **269,** 16,148–16,154.

# 14

# Synthetic Calcium-Binding Peptides

**Gary S. Shaw**

## 1. Introduction

The architecture of many calcium-binding proteins makes them exceptional candidates for synthetic peptide approaches. In particular, synthetic peptides have provided a wealth of insight into the calcium-binding properties and architecture of proteins in the EF-hand family of calcium-binding proteins. In this group of proteins, which includes members such as troponin C, calmodulin, and S100B, the calcium-binding sites are formed from contiguous stretches of about 30-residues ***(1)***. This sequence forms a helix-loop-helix structural motif whereby coordination of calcium occurs in a 12-residue loop region centered within the motif. The contiguous nature of the calcium-binding sites and their modular assembly is shown in **Fig. 1** for the muscle protein troponin C. Extensive use of the synthetic peptide approach has allowed a detailed examination of the importance of particular residues at the chelating positions to be addressed ***(2,3)***. Further, synthetic peptides from calcium-binding sites III and IV in troponin C have shown that the two-site domain is the integral building block of EF-hand calcium-binding proteins ***(4,5)***. Similar approaches using synthetic peptides have been used to study the assembly of calcium-binding sites in S100B ***(6)*** and calbindin $D_{28k}$ ***(7)***. Similar studies would have been relatively difficult to examine by other methods, such as site-directed mutagenesis. In addition, synthetic peptides have been used to study the calcium binding properties and three-dimensional structures of a variety of Gla-containing proteins ***(8)***.

Synthetic peptides offer several advantages for structure-function studies of calcium-binding proteins. First, they allow rapid production of a peptide typically in less than 1 wk, from idea to synthesis to purified peptide. Second, the efficiency of the method allows very high amounts of purified peptide, to be obtained, generally >25 mg per synthesis. Third, the method is sufficiently

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

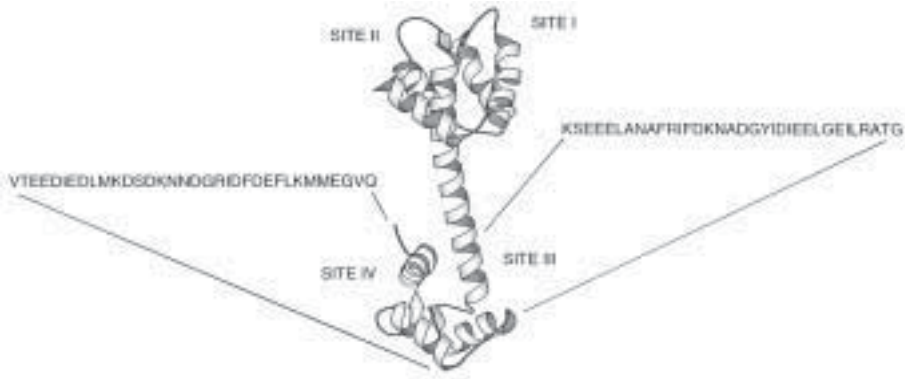


Fig. 1. Ribbon drawing of troponin C produced using Molscript ***(11)***. The figure shows the four helix-loop-helix motifs (sites I–IV) that are amenable to synthetic peptide studies. Also indicated are regions from sites III and IV corresponding to two 34-residue synthetic peptides which have been used to examine calcium affinity and assembly of EF hand calcium-binding proteins.

flexible to allow manipulation of the peptide synthesis *in situ*. This allows production of several similar sequences to be done from a single synthesis. This chapter will focus on the synthesis and purification of synthetic EF-hand calcium-binding peptides. The Boc method for synthesis will be described, although other methods are available. The utility of splitting of synthesis will be described using the example shown in **Fig. 2**.

## 2. Materials

### 2.1. Peptide Synthesis

1. Automated peptide synthesis instrument. Systems are available from a number of suppliers such as Applied Biosystems, Perceptive, and Tyler Research Corporation (Edmonton, Alberta, Canada).
2. Amino acids for Boc synthesis (Bachem, Philadelphia, PA). The following side-chain protecting groups are suggested: benzyl (Thr, Ser), p-toluene-sulphonyl (Arg), benzyl ester (Glu, Asp), 2-bromobenzyloxycarbonyl (Tyr), 2-chlorobenzyloxycarbonyl (Lys), 4-methoxybenzyl (Cys), and di-p-toluenesulphonate (His).
3. Solid-phase peptide support resin. A wide variety of resins are available. Typically one is chosen based on the characterization required at the *C*-terminus of the peptide or the method used for cleavage. Some popular resins include: benzhydrylamine (BHA), to obtain a neutral *C*-terminal amide and phenylacetoamidomethyl (PAM) ester to obtain a negatively charged carboxyl group. Frequently, resin may be purchased with the desired *C*-terminal residue already coupled to it.
4. HF Cleavage apparatus (Peninsula Laboritories, Belmont, CA).

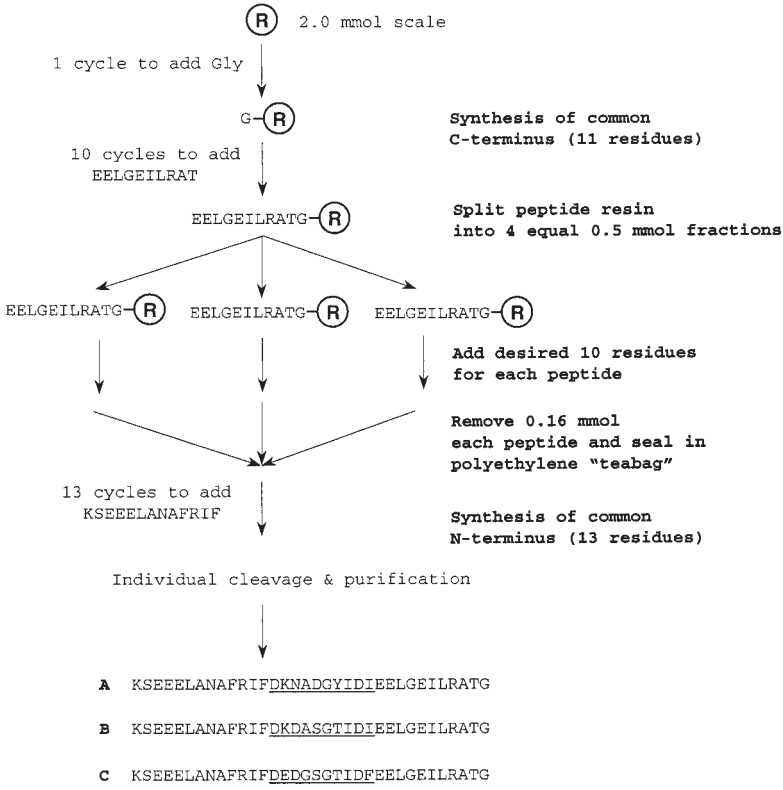


Fig. 2. Schematic diagram showing a potential synthetic route for production of three helix-loop-helix peptides having identical helix components, but different calcium-binding loops. Synthesis begins with the stepwise addition of the common 11 *C*-terminal amino acid residues to the resin. The peptide resin is then split and three independent syntheses conducted to add appropriate residues for the calcium-binding loops. In this case, three sequences representing 10 residues from the calcium-binding loops of troponin C are synthesized (underlined in the final sequences): site III **(A)**, site II **(B)**, and site II ligands with intervening residues from site III **(C)**. A portion of each of these synthesized peptide resins is then sealed in a polypropylene "teabag" and the synthesis is completed on the three "teabags" simultaneously.

## 2.2. HPLC Purification

1. A modern HPLC for purification consisting of a programmable solvent delivery system, two injector loops (50-µL and 5-mL capacities), variable wavelength detector, and recorder.
2. Two reversed-phase columns: analytical C8 column (220 × 4.6 mm id) and semipreparative C4 column (250 × 10 mm id)
3. HPLC grade water (*see* **Note 1**), $CH_3CN$, and trifluoroacetic acid (TFA).

### 2.3. Other

1. Chemicals for peptide synthesis: methylene chloride ($CH_2Cl_2$), diisopropylamine (DIEA), dimethyl formamide (DMF), dicyclohexylcarbodiimide (DCC).
2. Freeze-drying apparatus and appropriate lyophilization vessels.
3. Access to mass spectrometry facilities.

## 3. Methods

### 3.1. Peptide Synthesis

Synthesis of calcium-binding peptides is done using an automated peptide synthesizer, following the vendor's suggestions. This process involves the synthesis from the *C*-terminal residue to the *N*-terminal residue on a solid polymer resin ***(9)***. Usually, a resin is chosen that has the desired *C*-terminal residue chemically coupled. Each amino acid to be added is protected at its amino end by a tert-butoxycarbonyl (Boc) group and may have further side-chain protection (*see* **Subheading 2.1.**). Subsequent amino acids are added via (1) deprotection of the amino group on the resin attached peptide; (2) coupling of the next amino acid using DCC; and (3) recoupling of the same amino acid. At the completion of the synthesis the peptide is cleaved from the resin using liquid HF ***(10)***, which also removes side-chain protecting groups. Keep in mind that all peptides have unique properties and may require further experiments to determine the best reaction times, coupling protocols and number of couplings.

1. Weigh out an appropriate amount of starting resin based on the substitution ratio of amino acid/resin and the scale of the synthesis. For a substitution of 0.9 mmol amino acid/g resin, one would need 2.22 g of substituted resin for a synthesis on a 2.0 mmol scale.
2. Program synthesizer for the desired sequence. A sample protocol to couple each amino acid might be:
   a. Deprotection of Boc-group with aqueous TFA in $CH_2Cl_2$, followed by three $CH_2Cl_2$ washes;
   a. Neutralization with 10% DIEA in DMF.
   b. Amino acid first coupling.
   c. Three DMF washes.
   d. Second coupling.
   e. Five DMF washes.
3. Once synthesis is complete, the peptide may be *N*-terminal acetylated using acetic anhydride in DIEA (50:20) for every 1 part resin for 10 min.
4. Dry the synthesized peptide-resin over $P_2O_5$ overnight.
5. Transfer the dry peptide to a graduated HF-resistant reaction vessel (*see* **Note 2**). Calculate the amount of liquid HF required to cleave the peptide using 20 mL HF/g resin. Based on this volume of liquid HF, add anisole (10% v/v) and dithioethane (2% v/v) to the resin. Place the reaction vessel on the HF apparatus

and distill in the desired volume of HF keeping the reaction vessel at –4°C for 1 h. Remove the HF under vacuum using the apparatus.
6. Extract the peptide resin three times with 25 mL diethyl ether to remove organic impurties generated from side-chain deprotection.
7. Extract the resin three times with 25 mL glacial acetic acid to solubilize the peptide.
8. Dilute the acetic acid extract with at least 150 mL deionized water, transfer the solution to a lyophilization flask, and lyophilize until dry.

### 3.2. Multiple Sequences from One Synthesis

This procedure is useful for synthesis of calcium-binding peptides that have identical *N*- and *C*-termini (i.e., the helices), but different intermediate regions ("loops") as shown in **Fig. 2** ***(3)***.

1. Stop the synthesis after the identical *C*-terminal residues have been added.
2. Divide the resin into four equal portions corresponding to 0.5 mmol scale each.
3. For each portion of resin, carry out the next series of amino acid couplings according to the peptide sequence desired (10 residues in **Fig. 2**).
4. When complete, place a 0.16-mmol resin aliquot for each peptide in separate polypropylene bags (2.5 cm × 3.5 cm) and thermally seal the bag. Mark each bag with a distinguishing tag according to the sequence it contains (*see* **Note 3**).
5. Place the bags back in the synthesizer and carry out the remainder of the sequence.
6. Continue with **step 4** (*see* **Subheading 3.1.**).

### 3.3. Peptide Purification

#### 3.3.1. Sample Preparation

1. Dissolve 40–60 mg of lyophilized crude peptide in 3–4 mL of 50% aqueous trifloroacetic acid (TFA).
2. After a maximum of the peptide has dissolved, spin the sample in a benchtop centrifuge for 5 min. Transfer the clear supernatant to a clean tube.

#### 3.3.2. HPLC Preparation

1. Connect the analytical reversed-phase column to the instrument along with the 50 μL injection loop.
2. Equilibrate the HPLC system with the desired solvents. Popular solvents for peptide purification include 0.05% TFA/$H_2O$ (eluent A) and 0.05% TFA/$CH_3CN$ (eluent B).
3. Run one or two "blank" runs to ensure the column and system are clean. Use a flow rate of 1.0 mL/min and a linear gradient of 5% eluent B/min. Monitor the system at 210 nm.
4. Re-equilibrate the system with eluent A. Load 20 μL of the dissolved peptide solution and inject onto column. Run a linear gradient of 2% eluent B/min.
5. Record the spectrum to determine the retention time (i.e., % eluent B) the peptide mixture elutes. For the 34-residue calcium-binding peptides shown in **Fig. 2** this is typically near 35% B.

### 3.3.3. Purification

1. Connect the semipreparative reversed-phase column to the instrument along with the 5-mL injection loop.
2. Equilibrate the column with 0.05% eluent A at a flow rate of 5 mL/min until a constant baseline is observed.
3. Inject the entire dissolved peptide sample onto the column and wait approx 5–7 min for the void volume to elute.
4. Initiate the elution gradient. For peptides eluting near 35% eluent B, the following gradient is useful: time 0, 100% A; linear gradient 1% B for 28 min; linear gradient 0.1% B for 120 min; rapid linear gradient to 100% B (7.5% B/min for 8 min).
5. Using a programmable fraction collector, collect fractions at a rate of 1–2 mL/min during the 0.1% gradient period.

### 3.3.4. Peptide Analysis

In most cases the desired peptide will correspond to the largest peak in the chromatogram obtained in **Subheading 3.3.2.** Use this chromatogram as a reference for the following analysis.

1. Choose 3–4 fractions from the purification run and analyze these individually using the analytical column and the 50-µL injection loop.
2. For each sample mix 10–20 µL of each fraction with an equivalent amount of HPLC grade $H_2O$.
3. Equilibrate the column with eluent A.
4. Inject the sample and run a 2% B/min linear gradient. Monitor the run at 210 nm.
5. Choose samples that have similar retention times corresponding to that of the largest peak in the analytical chromatogram in **Subheading 3.3.2.** (*see* **Note 4**).
6. Choose two samples that appear to correspond to the same peptide. Mix appropriate amounts of each (10–20 µL) based on their relative peak heights obtained in the chromatograms in **step 5**.
7. Inject and monitor this sample. If the two peaks arose from the same peptide, a single peak will be obtained. Two peaks indicates the peptides are not the same species.
8. Repeat **step 7** as necessary to identify fraction tubes containing the same peptide.
9. Pool the fractions and analyze by mass spectrometry for the correct mass. A convincing mass spectrum will have a mass within ±1 atomic mass unit (amu) from the calculated peptide mass.

## 4. Notes

1. In our experience purifications done with HPLC grade $H_2O$ are more reproducible. Deionized $H_2O$ can be used and is less expensive; however, the $H_2O$ must be passed through a 0.22-µm filter before use.

2. Exercise extreme caution when using HF and the apparatus. A heavy HF-resistant apron and gloves should be worn at all times.
3. Mark the bags by cutting out a small, but distinctive, pattern along the edges but outside the thermal seam. This helps identify the peptide in each bag while protecting it from problems with labels drawn in ink or marker that will dissolve upon exposure to the chemicals in the reaction vessel.
4. Keep in mind that each sample injected will have some residual $CH_3CN$ in solution compared to the original unpurified sample. Therefore, the retention time for purified fractions will be shorter by a few percent compared to the original sample. When scanning a series of fractions for the desired purified peptide, the most diagnostic pattern to look for is a peak that increases in size to a maximum, then slowly decreases with all peaks having an identical retention time.

## References

1. Kretsinger, R. H. and Nockolds, C. E. (1973) Carp muscle calcium-binding protein. II. Structure determination and general description. *J. Biol. Chem.* **248,** 3313–3326.
2. Marsden, B. J., Hodges, R. S., and Sykes, B. D. (1988) $^1$H NMR studies of synthetic peptide analogues of calcium-binding site III of rabbit skeletal troponin C: effect on the lanthanum affinity of the interchange of aspartic acid and asparagine residues at the metal-ion co-ordinating positions. *Biochemistry* **27,** 4198–4206.
3. Shaw, G. S., Hodges, R. S., and Sykes, B. D. (1991) Probing the relationship between a-helix formation and calcium affinity in Troponin C: $^1$H NMR studies of calcium binding to synthetic and variant site III helix-loop-helix peptides. *Biochemistry* **30,** 8339–8347.
4. Shaw, G. S., Hodges, R. S., and Sykes, B. D. (1990) Calcium-induced peptide association to form an intact protein domain: $^1$H NMR structural evidence. *Science* **249,** 280–283.
5. Shaw, G. S., Findlay, W. A., Semchuk, P. D., Hodges, R. S., and Sykes, B. D. (1992) Specific formation of a heterodimeric two-site calcium-binding domain from synthetic peptides. *J. Am. Chem. Soc.* **114,** 6258–6259.
6. Donaldson, C., Barber, K. A., Kay, C. M., and Shaw, G. S. (1995) Human S100b protein: Formation of a tertramer from synthetic calcium-binding peptides. *Protein Sci.* **4,** 765–772.
7. Linse, S., Thulin, E., Gifford, L. K., Radzewsky, D., Hagan, J., Wilk, R. R., and Akerfeldt, K. S. (1997) Domain organization of calbindin D28k as determined from the association of six synthetic EF-hand fragments. *Protein Sci.* **6,** 2385–2396.
8. Rigby, A. C., Baleja, J. D., Furie, B. C., and Furie, B. (1997) Three-dimensional structure of a gamma-carboxyglutamic acid-containing conotoxin, conantokin G, from the marine snail Conus geographus: the metal-free conformer. *Biochemistry* **36,** 6906–6914.
9. Merrifield, R. B. (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154.

10. Hodges, R. S., Semchuk, P. D., Taneja, A. K., Kay, C. M., Parker, J. M., and Mant, C. T. (1988) Protein design using model synthetic peptides. *Peptide Res.* **1,** 19–30.
11. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24,** 946–950.

# 15

# Proteolytic Fragments of Calcium-Binding Proteins

**Richard D. Brokx and Hans J. Vogel**

## 1. Introduction

A major family of $Ca^{2+}$-binding proteins is the "EF-hand" superfamily ***(1)***, so-called because they all contain "EF-hand" helix-loop-helix $Ca^{2+}$-binding motifs. These motifs predominantly exist in pairs in these proteins, which is important for high-affinity, cooperative binding of $Ca^{2+}$ ions. Separation of individual helix-loop-helix domains through proteolysis provides an ideal starting point to asses the importance of having these domains in pairs. In addition to proteolytic methods, helix-loop-helix domains can be synthesized chemically through solid-phase peptide synthesis ***(2)***. This permits total freedom in choosing mutation sites and the location of the start and end of the polypeptide chains, but can be cost-prohibitive because of the size (approx 35 residues) of the peptides needed. Whatever the route they obtain, isolated EF-hands are interesting models of $Ca^{2+}$-binding proteins. Often, these isolated motifs, such as thrombic fragements of calmodulin (CaM) ***(3)*** and synthetic troponin-C peptides ***(2,4,5)*** dimerize in vitro to form native-like structures.

Moreover, most EF-hand proteins, including regulatory proteins, such as CaM and troponin C, and $Ca^{2+}$ "buffering" proteins, such as parvalbumin and the calbindins, have more than one pair of EF-hands contained within one polypeptide molecule. Here, individual pairs of helix-loop-helix domains can also be created through proteolysis. With these larger fragments, the contribution of individual lobes to the function of regulatory $Ca^{2+}$-binding proteins, or the importance of higher-order domain organization in $Ca^{2+}$-buffering proteins, can be evaluated. Once the various possible fragments of $Ca^{2+}$-binding proteins are created, researchers can employ various spectroscopic techniques, activity assays, or other structural biology tools to examine the properties of these isolated domains. Techniques discussed in this volume that could be used

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

**Table 1**
**An Overview of the Fragments of Various $Ca^{2+}$-Binding Proteins Studied in the Literature, and the Methods by which they were Produced[a]**

| Protein | Fragments | Method | Refs. |
|---|---|---|---|
| Calmodulin | TR1C (1–75), TR2C (78–148) | Recombinant expression | ***15–17*** |
| | TR1C (1–77), TR2C (78–148) | Tryptic digest | ***16–22*** |
| | TM1 (1–106), TM2 (107–148) | Thrombic digest | ***14,18*** |
| | 1–37, 38–148 | Thrombic digest | ***23*** |
| Troponin-C | 9–84, 89–159 | Tryptic digest | ***13,19*** |
| | TH1 (1–120), TH2 (121–159) TH3 (1–100) | Thrombic digest | ***14,24,25*** |
| | SCIII (93–126), SCIV (129–162) | Peptide synthesis | ***2,4,5*** |
| Calbindin $D_{9k}$ | 1–43, 44–75 | Mutagenesis/ CNBr cleavage[b] | ***26*** |
| Calbindin $D_{28k}$ | EF1, EF2, EF3 EF4, EF5, EF6 | Peptide synthesis | ***27*** |
| SCBP[c] | 1–80, 90–174 | Tryptic digest | ***28*** |
| Parvalbumin | 76–108 | Clostripain digest | ***29*** |
| α-lactalbumin | 1–90 | CNBr cleavage[b] | ***30*** |

[a] With the exception of α-lactablbumin, all are EF-hand proteins.

[b] CNBr cleavage: cyanogen bromide chemically cleaves after methionine residues in proteins, which may be introduced through site-specific mutagenesis or already present in the amino acid sequence.

[c] SCBP: sarcoplasmic calcium-binding protein.

to study proteolytic fragments include optical spectroscopy ***(6)***, Fourier-transform infrared spectroscopy ***(7)***, fluorescence spectroscopy ***(8)***, and NMR of various nuclei ***(9–11)***.

This chapter gives an overview of the cleavage techniques used for various calcium-binding proteins, and gives detailed methods for cleavage of CaM or troponin-C(TnC) with trypsin and thrombin. Trypsin cleaves $Ca^{2+}$–CaM ***(12)*** or $Ca^{2+}$–TnC ***(13)*** largely in the central linker region to yield two half-molecules. Thrombin, on the other hand ***(14)***, can cleave CaM or TnC between the two $Ca^{2+}$-binding sites in the *C*-terminal lobe of the proteins. *See* **Table 1** for a list of $Ca^{2+}$-binding proteins and their fragments, which have been examined in the literature. Detailed procedures follow, which we have used to cleave CaM

with trypsin or thrombin in our laboratory. The procedures are virtually identical for the highly homologous protein TnC.

## 2. Materials

### 2.1. Proteolytic Digestion

1. Lyophilized protein sample. Approximately 100 mg of CaM or TnC is needed for preparation of proteolytic fragments. It need not be totally pure of other polypeptides because the digest will be repurified anyway, but it should be free of salts, buffers, and other contaminants.
2. Buffers/reagents. Stock solutions that should be on hand include:
   a. 1 *M* Tris-HCl. Can be kept for months at 4°C. Generally, a pH 7.5 stock solution is kept on hand, but the pH can be altered when using it to make other buffers.
   b. 0.5 *M* Ethylenediamine tetraacetic acid (EDTA) or ethylene glycol *bis*-(β-aminoethyl ether)-*N*,*N*,*N*',*N*'-tetraacetic acid (EGTA), pH 8.0. Dissolve the free acid or the disodium salt and adjust the pH to 8.0 with solid NaOH and 12 *M* NaOH. Either solution is suitable; EDTA is a more generally used laboratory reagent, but EGTA is more appropriate because it is a more specific chelator for $Ca^{2+}$ ions. These solutions can be kept indefinitely at room temperature.
   c. 0.1 *M* Dithiothreitol (DTT). Store frozen in small aliquots.
   d. 50 m*M* Ammonium bicarbonate ($NH_4HCO_3$). Store at 4°C. The pH of this solution at 4°C is 7.9.
3. Incubator (37°C). One used for growing bacterial plates is suitable.
4. UV/visible spectrophotometer. Used for checking protein concentrations in column fractions. *See* **Fig. 1** for examples of the UV absorption spectra of CaM and its proteolytic fragments.
5. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) apparatus. Used for checking purity of protein fractions. A 20% acrylamide gel is suitable for fragments of these sizes. *See* **Fig. 2** for an example of an SDS-PAGE gel of CaM and its proteolytic fragments.
6. Chelex-100 matrix (Bio-Rad). Used for removal of metal ions from buffers and protein samples.

### 2.2. Tryptic Digestion

Prepare all buffers fresh.

1. 1 m*M* HCl.
2. Buffer A: 50 m*M* $NH_4HCO_3$, 50 m*M* NaCl, pH 7.9.
3. Buffer B: 50 m*M* Tris-HCl, 1 m*M* $CaCl_2$, pH 7.5.
4. Buffer C: 2 m*M* Tris-HCl, 1 m*M* $CaCl_2$, pH 7.5.
5. Double-distilled (dd) $H_2O$, pH 7.5. Adjust pH very carefully with small amounts of NaOH and HCl.

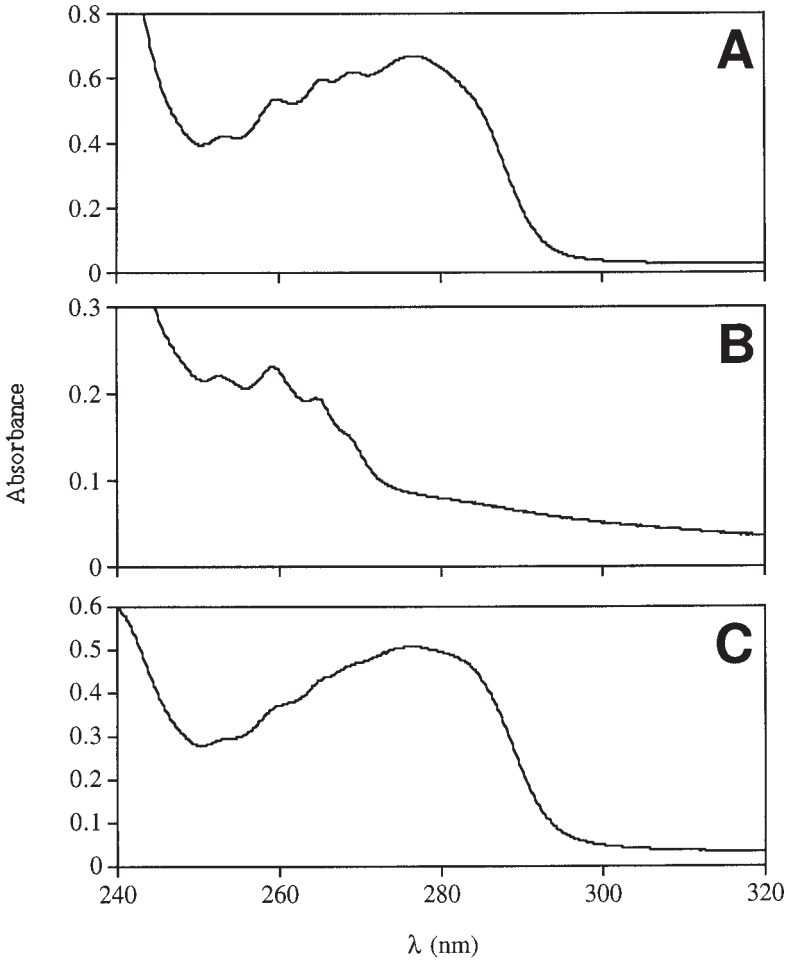


Fig. 1. UV absorption spectra of calmodulin and its proteolytic fragements. **(A)** calmodulin. **(B)** TR1C. **(C)** TR2C.

6. Trypsin (Sigma), from bovine pancreas; 10,000–13,000 BAEE U/mg protein. Lyophilized. Should be purchased TPCK-treated to reduce chymotrypsin activity (< 0.1 BTEE U/mg protein). Dissolve 5 mg in 1 mL 1 m*M* HCl. Prepare fresh.
7. Soybean trypsin inhibitor (STI) (Sigma). Lyophilized. 1 mg of STI inhibits 1–3 mg of trypsin with an activity of approx 10,000 BAEE units/mg protein. Dissolve 5 mg in 1 mL 1 m*M* HCl. Prepare fresh.
8. Sephadex G-50 (Pharmacia) (100 cm × 1.5 cm) equilibrated with buffer A.
9. Phenyl-Sepharose (Pharmacia) (approx 30 mL) equilibrated with buffer B.

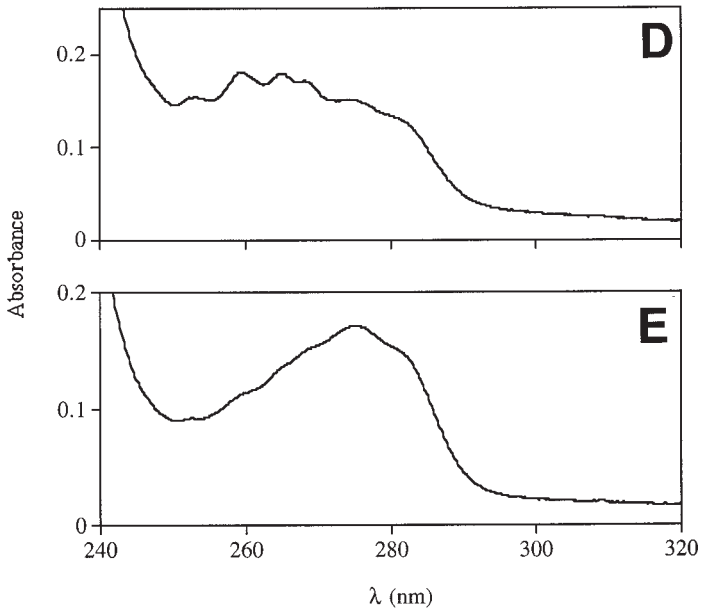


Fig. 1. (*continued*) **(D)** TM1. **(E)** TM2.

### 2.3. Thrombic Digestion

1. 100 m*M* Tris-HCl, pH 8.5. Can be stored for months at 4°C. Chelex-treated 50 m*M* $NH_4HCO_3$. To remove all metal ions, stir buffer using a magnetic stirrer with approx 5 mL Chelex-100 resin per liter of buffer for about 5 min, then filter off the resin. Can be stored for weeks at 4°C.
2. Thrombin (Sigma), from bovine plasma. Should be purchased free of all other clotting factors and proteolytic activities. Reconstitute according to the package directions or dissolve in an appropriate buffer (e.g., sodium citrate) to a concentration of 1 U/mL. This enzyme is intolerant to freeze-thawing; dispense into aliquots of 50–100 U and store at –70°C.
3. Sephadex G-50 (Pharmacia) (100 cm × 1.5 cm) equilibrated with chelexed 50 m*M* $NH_4HCO_3$.

## 3. Methods

### 3.1. Tryptic Digestion of Calmodulin

1. Pre-equilibrate a Sephadex G-50 column (100 × 1.5 cm) with buffer A at approx 0.35 mL/min overnight.
2. Dissolve 100 mg CaM in 4 mL freshly prepared buffer A (to approx 1.5 m*M* CaM concentration). Add five equivalents $CaCl_2$ (to approx 7.5 m*M*; add 15 mL of a 2 *M* stock). Take a sample for SDS-PAGE.

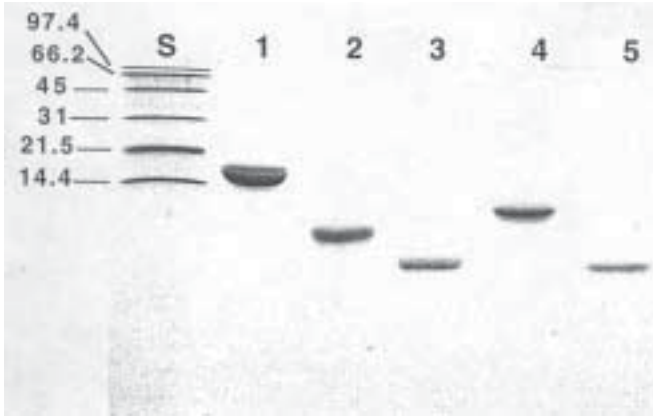


Fig. 2. 20% SDS-PAGE gel of calmodulin and its proteolytic fragments. Lane S: protein standards (molecular weight indicated in kDa), lane 1: calmodulin, lane 2: TR1C, lane 3: TR2C, lane 4: TM1, lane 5: TM2.

3. Add 200 mL of the trypsin stock (to 250 mg/mL trypsin) and incubate for 40 min at 37°C.
4. After the 40 min, add 400 mL of the soybean trypsin inhibitor stock (to 500 mg/mL STI), and cool the mixture on ice. Take a sample for SDS-PAGE (*see* **Fig. 2**).
5. Quickly apply the digest to the G50 column by the following procedure:
   a. First, disconnect the column from the pump and let the excess buffer run into the top of the column until only a very tiny amount of buffer covers the top of the matrix.
   b. Then very gently apply the digest solution and let it run into the column in the same manner.
   c. Repeat this again a few times with small amounts (1 mL) of buffer A until the sample is washed completely into the column.
   d. Then gently apply some buffer A on top of the column and reconnect the column to the pump. Run the column with buffer A at approx 0.35 mL/min. Collect 15 min fractions.
6. Measure $A_{280nm}$ of the column fractions. There will basically be two peaks; the first will be undigested protein and trypsin as well as its inhibitor; the second (about 50 mL later) will be the two tryptic fragments, TR1C and TR2C. Collect the second peak and assess purity by SDS-PAGE (*see* **Fig. 2**).
7. Equilibrate a phenyl-Sepharose column with buffer B. Run at 2 mL/min for approx 2 h. The phenyl-Sepharose matrix is very stable; it may be quickly equilibrated by removing the storage solution and then resuspending in buffer B.
8. Apply the pooled fractions onto the phenyl-Sepharose column and run at approx 2 mL/min. CaM and its tryptic fragments will bind to this matrix due to the

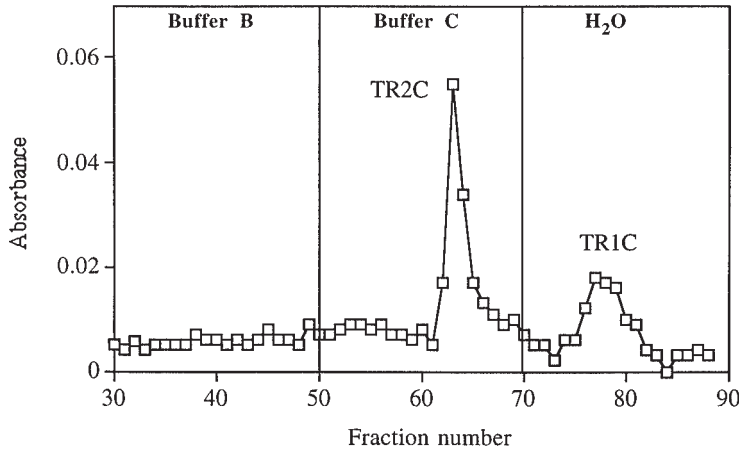


Fig. 3. Elution profile of tryptic fragments of calmodulin from phenyl-Sepharose. The fractions at which the buffer was switched are indicated with vertical lines. Absorbance values are at 280 nm until fraction 70 after which they are at 258 nm (TR1C only has phenylalanine chromophores).

$Ca^{2+}$-dependent exposure of hydrophobic patches, one in each lobe ***(19)***. TR1C binds to this column more strongly than TR2C, and thus they can be separated.

9. Elute with:

   1 column volume (approx 30 mL) buffer B, then
   2 column volumes (approx 60 mL) buffer C, then
   3 column volumes (approx 100 mL) $ddH_2O$, pH 7.5.

   Collect 2 min fractions throughout, and check them for protein by monitoring absorbance on a UV spectrophotometer (*see* **Fig. 3**). The TR2C fragment elutes with buffer C (at about 50–60 mL); it can be detected by monitoring $A_{280nm}$ because of the presence of two tyrosine residues. TR1C elutes with $ddH_2O$ (at about 120–140 mL); it has no tyrosine residues, so it must be detected by monitoring $A_{258nm}$. The yield of TR1C is generally lower than that of TR2C (*see* **Note 1**). Assess purity of the fractions by SDS-PAGE, amino acid analysis, or other methods (*see* **Note 2**).
10. Quantitation. Ideally TR1C and TR2C should be quantitated by quantitative amino acid analysis. However, we have calculated molar extinction coefficients of $e_{258nm}$ = 1073 *M*/cm for TR1C and $e_{276nm}$ = 2666 *M*/cm for TR2C ***(17)***. *See* **Fig. 1** for UV absorption spectra of TR1C and TR2C.

### *3.2. Thrombic Digestion of Calmodulin*

1. The CaM used for thrombic digestion (*see* **Note 3**) should ideally be free of all metal ions; dissolve the desired amount of protein in 50 m*M* $NH_4HCO_3$ and pass

it through an approx 5 mL column of Chelex-100 resin. Collect the eluted protein and lyophilize.
2. Pre-equilibrate a Sephadex G50 (Pharmacia) (100 cm × 1.5 cm) with chelexed 50 m*M* $NH_4HCO_3$ at approx 0.35 mL/min overnight.
3. Dissolve 100 mg apo-CaM in digest solution (final concentration 50 m*M* Tris-HCl, pH 8.5, 1 m*M* DTT, 5 m*M* EGTA):
   2.5 mL 100 m*M* Tris-HCl, pH 8.5;
   2.3 mL $H_2O$;
   50 mL 0.1 *M* DTT;
   50 mL 0.5 *M* EGTA.
   Take a sample for SDS-PAGE.
4. Add thrombin (1 U/mg CaM; 100 mL of a 1 U/mL solution) and incubate at 37°C for 90 min. After the digestion, cool the mixture in an ice bath. Take a sample for SDS-PAGE (*see* **Note 4**).
5. Immediately load the digest solution on the G50 column with the same method as was used for the tryptic fragments. Run with chelexed 50 m*M* $NH_4HCO_3$ at approx 0.35 mL/min. Collect 15-min fractions. The first proteins should elute from the column in about 6 h. Thrombin elutes first and is removed easily; then comes intact CaM, then TM1, and finally, TM2. Monitor fractions by absorbance at 280 nm (both TM1 and TM2 have a tyrosine residue) and check their purity by SDS-PAGE (*see* **Fig. 2**).
6. In practice, we usually collect the fractions in four pools: 0: intact CaM (if any) and TM1, can be digested again in the next repeat procedure; 1: pure TM1; 2: TM1 and TM2, can be repurified in the next repeat procedure or by phenyl-Sepharose chromatography (*see* **Note 5**); and 3: pure TM2. Assess the purity of these pools by SDS-PAGE (*see* **Fig. 2**).
7. Quantitation. Ideally, TM1 and TM2 should be quantified by quantitative amino acid analysis. However, we have calculated molar extinction coefficients of $e_{265nm}$ = 2360 *M*/cm for TM1 and $e_{275nm}$ = 1860 *M*/cm for TM2 ***(3)***. *See* **Fig. 1** for UV absorption spectra of TM1 and TM2.

## 4. Notes

1. The yield of TR1C by tryptic digestion of CaM is lower than that of TR2C. However, in our hands, we are able to recombinantly express and purify TR1C much more easily than TR2C ***(17)***. Thus, the two techniques complement each other; TR1C is expressed in *Escherichia coli* cells, whereas TR2C is produced from tryptic digestion of CaM. Recombinantly expressed TR1C can be purified in the same manner as intact CaM ***(19,31)***.
2. TR1C and TR2C produced by this tryptic digestion are often heterogeneous because of cleavage of CaM by trypsin at several points (Arg74, Lys75, Lys77; ***20***) in the central linker region of the molecule. However, the majority should be the fragments 1–77 and 78–148 ***(17,22)***.
3. Often, bacterially expressed CaMs produced in our laboratory have the three *C*-terminal residues removed (unpublished observations), which, in turn, can limit

the yield of TM2 in this procedure. To combat this, ensure that after expressing CaM, bacterial cell pellets are washed thorougly in an EDTA-containing buffer, such as 50 m*M* Tris-HCl, pH 7.5, 2 m*M* EDTA, 1 m*M* DTT, 0.85% NaCl, before storage at –20°C.

4. If the thrombic digests is allowed to proceed too long, or if $Ca^{2+}$ is not completely removed ***(23)***, an additional fragment (1–37 of CaM) can be produced. This fragment runs very close to TM2 (108–148) on an SDS-PAGE gel and may not be seen. As well, it copurifies with TM2 on a G50 column. Use anion-exchange chromatography (DEAE-Sephadex or other) to separate TM2 from this fragment because of their difference in ionic charges. Again, however, all of this should not be necessary if the thrombic digest of CaM is not allowed to proceed too long.
5. In practice, it is possible to purify the thrombic fragments of CaM by phenyl-Sepharose chromatography in a manner analogous to TR1C and TR2C ***(19)***. However, it is difficult to separate TM1 from CaM by this method and, moreover, gel-filtration can be performed in a volatile buffer (50 m*M* $NH_4HCO_3$) that enables direct lyophilization of the purified fragments.

## Acknowledgments

This research was funded by a general operating grant from the Medical Research Council of Canada (MRC). R. D. Brokx is the recipient of studentships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Alberta Heritage Foundation for Medical Research (AHFMR). H. J. Vogel is the recipient of a scientist award from AHFMR.

## References

1. Strynadka, N. C. J. and James, M. N. G. (1989) Structures of the helix-loop-helix calcium binding proteins. *Annu. Rev. Biochem.* **58,** 951–998.
2. Shaw, G. S. (2001) Synthetic calcium-binding peptides, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey, pp. 175–182.
3. Brokx, R. D. and Vogel, H. J. (2000) Peptide and metal ion dependent association of isolated helix-loop-helix calcium binding domains: studies of thrombic fragments of calmodulin. *Protein Sci.* **9,** 964–975.
4. Shaw, G. S., Hodges, R. S., and Sykes, B. D. (1990) Calcium-induced peptide association to form an intact protein domain: $^1H$ NMR structural evidence. *Science* **249,** 280–283.
5. Shaw, G. S. and Sykes, B. D. (1996) NMR structure of a synthetic troponin C heterodimeric domain. *Biochemistry* **35,** 7429–7438.
6. Martin, S. R. and Bailey, P. M. (2001) Absorption and circular dichroism spectroscopy, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey, pp. 43–56.
7. Fabian, H. and Vogel, H. J. (2001) Fourier-transform infrared spectroscopy, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey, pp. 57–74.

8. Weljie, A. M. and Vogel, H. J. (2001) Steady state fluorescence spectroscopy, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey, pp. 75–88.
9. Clarke, T. E. and Vogel, H. J. (2001) Cadmium-113 and lead-207 NMR spectroscopic studies of calcium-binding proteins, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey, pp. 205–216.
10. Drakenberg, T. (2001) Calcium-43 NMR spectroscopy of calcium-binding proteins, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey, pp. 217–230.
11. Li, M. X., Corson, D. C., and Sykes, B. D. (2001) Structure determination by NMR 1 — Isotope labeling, in Calcium-Binding Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J. ed.), Humana Press, Totowa, New Jersey.
12. Walsh, M., Stevens, F. C., Kuznicki, J., and Drabikowski, W. (1977) Characterization of tryptic fragments obtained from bovine brain protein modulator of cyclic nucleotide phosphodiesterase. *J. Biol. Chem.* **252,** 7440–7443.
13. Grabarek, Z., Drabikowski, W., Vinokurov, L., and Lu, R. C. (1981) Digestion of troponin-C with trypsin in the presence and absence of $Ca^{2+}$. Identification of cleavage points. *Biochim. Biophys. Acta* **671,** 227–233.
14. Wall, C. M., Grand, R. J. A., and Perry, S. V. (1981) Biological activities of the peptides obtained by digestion of troponin C and calmodulin with thrombin. *Biochem. J.* **195,** 307–316.
15. Finn, B. E., Evenäs, J., Drakenberg, T., Waltho, J. P., Thulin, E., and Forsén, S. (1995) Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* **2,** 777–783.
16. Bentrop, D., Bertini, I., Cremioni, M. A., Forsén, S., Luchinat, C., and Malmendal, A. (1997) Solution structure of the paramagnetic complex of the N-terminal domain of calmodulin with two $Ce^{2+}$ ions by $^1H$ NMR. *Biochemistry* **36,** 11,605–11,618.
17. Yuan, T., Ouyang, H., and Vogel, H. J. (1999) Surface exposure of the methionine side chains of calmodulin in solution. A nitroxide spin label and two-dimensional NMR study. *J. Biol. Chem.* **274,** 8411–8420.
18. Andersson, A., Forsén, S., Thulin, E., and Vogel, H. J. (1983) Cadmium-113 nuclear magnetic resonance studies of proteolytic fragments of calmodulin: assignment of strong and weak cation binding sites. *Biochemistry* **22,** 2309–2313.
19. Vogel, H. J., Lindahl, L., and Thulin, E. (1983) Calcium-dependent hydrophobic interaction chromatography of calmodulin, troponin C, and their proteolytic fragments. *FEBS Lett.* **157,** 241–246.
20. Thulin, E., Andersson, A., Drakenberg, T., Forsén, S., and Vogel, H. J. (1984) Metal ion and drug binding to proteolytic fragments of calmodulin: proteolytic, cadmium-113, and proton nuclear magnetic resonance studies. *Biochemistry* **23,** 1862–1870.
21. Linse, S., Helmersson, A., and Forsén, S. (1991) Calcium binding to calmodulin and its globular domains. *J. Biol. Chem.* **266,** 8050–8054.

22. Fabian, H., Yuan, T., Vogel, H. J., and Mantsch, H. H. (1996) Comparative analysis of the amino- and carboxy-terminal domains of calmodulin by Fourier transform infrared spectroscopy. *Eur. Biophys. J.* **24,** 195–201.
23. Shea, M. A., Verhoeven, A. S., and Pedigo, S. (1996) Calcium-induced interactions of calmodulin domains revealed by quantitative thrombin footprinting of Arg37 and Arg106. *Biochemistry* **35,** 2943–2957.
24. Kay, L. E., Forman-Kay, J. D., McCubbin, W. D., and Kay, C. M. (1991) Solution structure of a polypeptide dimer comprising the fourth $Ca^{2+}$-binding site of troponin C by nuclear magnetic resonance spectroscopy. *Biochemistry* **30,** 4323–4333.
25. Francois, J.-M., Sheng, Z., Szczesna, D., and Potter, J. D. (1995) The functional role of the domains of troponin-C investigated with thrombin fragments of troponin-C reconstituted into skinned muscle fibers. *J. Biol. Chem.* **270,** 27–34.
26. Finn, B. E., Kördel, J., Thulin, E., Sellers, P., and Forsén, S. (1992) Dissection of calbindin $D_{9k}$ into two $Ca^{2+}$-binding subdomains by a combination of mutagenesis and chemical cleavage. *FEBS Lett.* **298,** 211–214.
27. Åkerfeldt, K. S., Coyne, A. N., Wilk, R. R., Thulin, E., and Linse S. (1996) $Ca^{2+}$-binding stoichiometry of calbindin $D_{28k}$ as assessed by spectroscopic analyses of synthetic peptide fragments. *Biochemistry* **35,** 3662–3669.
28. Durussel, I., Luan-Rilliet, Y., Petrova, T., Takagi, T., and Cox, J. A. (1993) Cation binding and conformation of tryptic fragments of Nereis sarcoplasmic calcium-binding protein: calcium-induced homo- and heterodimerization. *Biochemistry* **32,** 2394–2400.
29. Revett, S. P., King, G., Shabanowitz, J., Hunt, D. F., Hartman, K. L., Laue, T. M., and Nelson, D. J. (1997) Characterization of a helix-loop-helix (EF-hand) motif of silver hake parvalbumin isoform B. *Protein Sci.* **6,** 2397–2408.
30. Berliner, L. J., Meinholtz, D. C., Hirai, Y., Musci, G., and Thompson, M. P. (1991) Functional implications resulting from disruption of the calcium binding loop in bovine α-lactalbumin. *J. Dairy Sci.* **74,** 2394–2402.
31. Putkey, J. A., Slaughter, G. R., and Means, A. R. (1985) Bacterial expression and characterization of proteins derived from the chicken calmodulin cDNA and a calmodulin processed gene. *J. Biol. Chem.* **260,** 4704–4712.

# 16

# Electron Magnetic Resonance Studies of Calcium-Binding Proteins

**Lawrence J. Berliner**

## 1. Introduction

This chapter, which focuses on electron spin resonance (ESR) or electron paramagnetic resonance (EPR) studies of calcium-binding proteins, is limited mainly to paramagnetic ions that mimic calcium (Mn[II] and the lanthanides) and to spin labeling of protein amino acid residues. Although the "proper" name of the technique should be electron magnetic resonance (EMR) to correlate with the NMR technique, this term has never been internationally adopted. Hence, the two terms ESR and EPR, which are also equivalent and are used ubiquitously and interchangeably in this chapter. Most of the calcium-binding proteins, with the exception of α-lactalbumin and one or two other proteins whose 3D structures are known are of the EF-hand class. Most of the proteins in this class are very similar in how they coordinate the bound metal ion (apparently all octahedral). Below, we review some of these results with proteins such as calmodulin, parvalbumin, troponin C, which are contrasted with α-lactalbumin, a protein characterized by a distorted trigonal bipyramidal coordination sphere.

### *1.1. Paramagnetic Metals*

The EPR of paramagnetic metal centers reflects both the geometry and electronic (ligand field) environment, which may be correlated with results from model systems. Most commercial EPR spectrometers are limited to X-band (9.5 GHz, 3.5 kG) although a few exist at Q-band (35 GHz). The spectral resolution for metal ions at X-band is usually not that good, making detailed simulation analyses difficult. With the relatively recent advent of high-field EPR (HF-EPR) at W-band (95 GHz) and above, the future promises to yield much

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

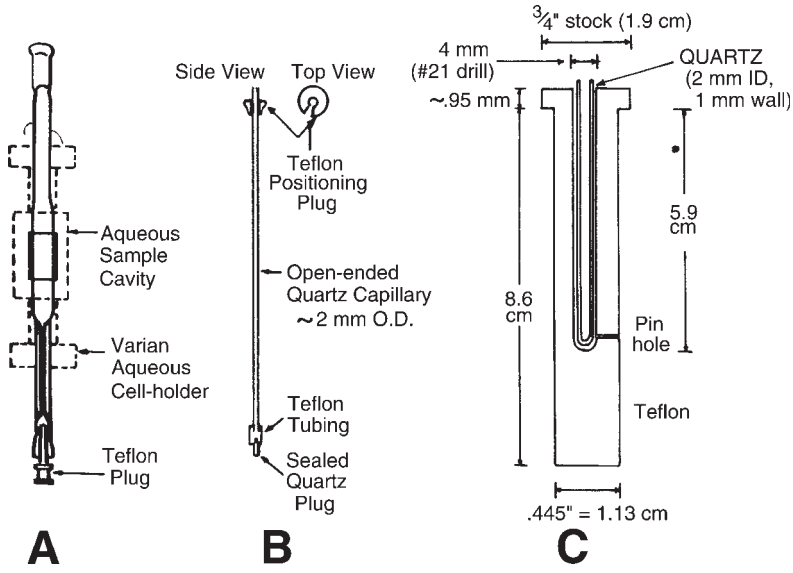


Fig. 1. Aqueous X-band ESR cell designs. **(A)** Standard quartz flat cell — this is probably the highest sensitivity cell design; cell positioning reproducibility is quite variable. **(B)** Capillary sample tube. The tube is open at both ends to facilitate filling and cleaning. The Teflon holder at the top ensures reproducible positioning of the tube in the cavity and is machined to fit the cavity port opening. The "plug" at the bottom is either Teflon or tygon tubing. This is designed from small sample volumes and is excellent for titration studies. **(C)** ESR "dispopipet" sample holder for routine measurements. This hold fits snugly into a rectangle and accepts sealed 9-in-long disposable Pasteur pipets as sample tubes. The top end is stoppered with a size 00 stopper or cork. Reproduced with permission from **ref. *3***.

more detailed information. At this juncture, however, most of the data is restricted to X-band or 35 GHz.

### *1.2. Spin Labeling*

The spin-label technique takes advantage of the motionally sensitive, highly chemical stable nitroxyl group (sometimes misnamed as nitroxides). If the molecule is intended for covalent modification of a macromolecule, it is termed spin label; if it is intended for a noncovalent probe, it is called a spin probe ***(1–3)***. Spin labeling requires the covalent attachment of a nitroxyl compound to a specific amino acid residue with a stoichiometry of unity. This is generally difficult for any residue but cysteine ***(4)***. Because most calcium-binding proteins generally do not contain free-cysteine residues, the next most specific residue is methionine. One may also combine the two approaches by taking advantage of

the spin–spin interaction between the paramagnetic metal center and the electron at the nitroxyl group in order to measure intramolecular distances ***(5)***. Last, when labeling other residues, such as terminal α-amino groups or methionine thioether linkages, a careful ion-exchange separation is necessary ***(6)***.

## 2. Materials

1. A commercial EPR spectrometer. Current major commercial sources of X-band (9.5 GHz) machines are Bruker Instruments and JEOL. Varian Associates eliminated ESR from their analytical instrument portfolio about 15–20 yr ago; however, several hundred instruments (E-series and Century Series) are still in service as the construction and quality was the "Model-T" and "Cadillac" of EPR spectrometers.
2. For selected studies, Q-band (35 GHz) is desirable as well as very-high-field W-band (95GHz), both of which are available from Bruker.
3. Samples are placed in either 100–250 μL quartz aqueous flat cells (X-band) or quartz cylindrical EPR sample tubes (inner diameter [id] ≤ 3 mm) or microcapillaries for Q-band (μL) or W-band (nL). **Figure 1** depicts some of these cells for X-band. The standard quartz flat cell gives the highest sensitivity, but cell-positioning reproducibility is difficult. The capillary sample tube is designed for small volume titration studies. The ESR "dispopipette" sample holder is for routine measurements on expendable samples.
4. Low-temperature spectra are frequently desired for metal ion studies. The sample cells are immersed in a liquid nitrogen dewar insert that fits into the EPR resonant cavity. Experiments at liquid helium temperatures require special (Oxford Instruments) helium transfer dewars.

## 3. Methods

### 3.1. Instrument Preparation

1. Before turning on the instrument, the cooling water must be circulating through the magnet, magnet power supply, and klystron/microwave bridge. All instruments have a thermally activated shutoff relay if the water circulation is not sufficient.
2. Because electromagnetic fields and klystron (or Gunn diode) frequency sources stabilize quite rapidly, an instrument warm up in the *STANDBY* mode of 15–30 min is sufficient. Obviously, a super-conducting magnet system such as that utilized at W-band, requires longer stabilization.
3. Tuning (matching) the sample is a critical, more sensitive operation. The sample tube is placed in the cavity: the flat cells utilize special clips, which screw onto the upper and lower parts of the $TE_{102}$ or $TM_{011}$ cavity. Capillary cells or quartz tubes require special, nonmagnetic fittings to align them properly (and reversibly for accurate concentration measurements). The sample is placed in the cavity in the *TUNE* mode; a cavity field sweep allows one to find and optimize the "Q-dip" for tuning by turning the frequency adjustment. Sample position is adjusted both mechanically and finally by the cavity iris adjustment screw. A motorized robot

in the Bruker instrument does the latter; however, if the screw is turned in too far, the iris is damaged. The instrument is switched to the *OPERATE* mode and the iris is fine-tuned to a detector leakage current of 200–250 μA.

4. Spectra are measured by sweeping the applied magnetic field typically over a 4–20-min scan period, depending on the sweep width (100 gauss [G] to kilogauss [kG]).

### 3.2. Sample Preparation

1. Ultrapure metal salts (>99.9%) should be used. Where the protein concentration is limited, precautions should be taken to eliminate paramagnetic impurities in other buffers, and so on. This is accomplished by a careful pass of the final buffer down a Chelex-100 column (but before adding calcium!).
2. Maintaining pH, by the choice of appropriate buffer concentrations, is crucial as always.
3. Purified dry protein is weighed and dissolved in the buffer of choice. As aforementioned, precautions should be taken to eliminate paramagnetic impurities in buffers and solvents. It may be necessary to dialyze the protein against EDTA or another chelator; alternatively, one can sometimes pass it down a Chelex-100 or Tris-EDTA column before adding calcium ***(8)***. Concentration can be verified by measuring the absorbency at 280 nm. It is usually common to add an aliquot of metal ion requisite stoichiometry and allow the sample to equilibrate for a few minutes.
4. For spin-labeling experiments, excess spin label is added to a solution of protein and allowed to react at 4°C with stirring for 2–4 h followed by exhaustive overnight dialysis. A sample of the last dialysate is checked by EPR for the presence of free, unremoved label.
5. The spin labels employed fall into two quite general classes:
   a. Those which are reactive with nucleophilic side chains principally those which alkylate amino groups (Lys) α-amino groups, (Cys) thioester groups (Met), and hydroxyl groups (Thr). Limited examples exist for modifying this imidazole and carboxyl group (Glu, Asp, α-carboxyl). Some of these labels, particularly the maleimide and iodoacetamido nitroxyls, are available from Sigma Chemical Company with varying "tether" lengths between the functional group and the nitroxyl ring. **Figure 2** shows some representative structures.
   b. Site-specific spin labels: to date, the only highly specific labels are based on the alkyl-thiolsulfonate reagents or nitroxyl disulfide biradicals, which, after reaction with the protein, result in the disulfide interchange product: Cys-S-S-spin label ***(9)***. For calcium-binding proteins, the calmodulin species with a single unique Cys have been the most successful. However, with the advent of molecular biological techniques, one can selectively substitute any amino acid residue in a protein by Cys. The most successful approach recently has been site-directed spin labeling where a residue is systematically substituted by Cys in the amino acid sequence (assuming that it folds correctly), followed by labeling with one of the thiol-specific spin labels shown in **Fig. 2**.

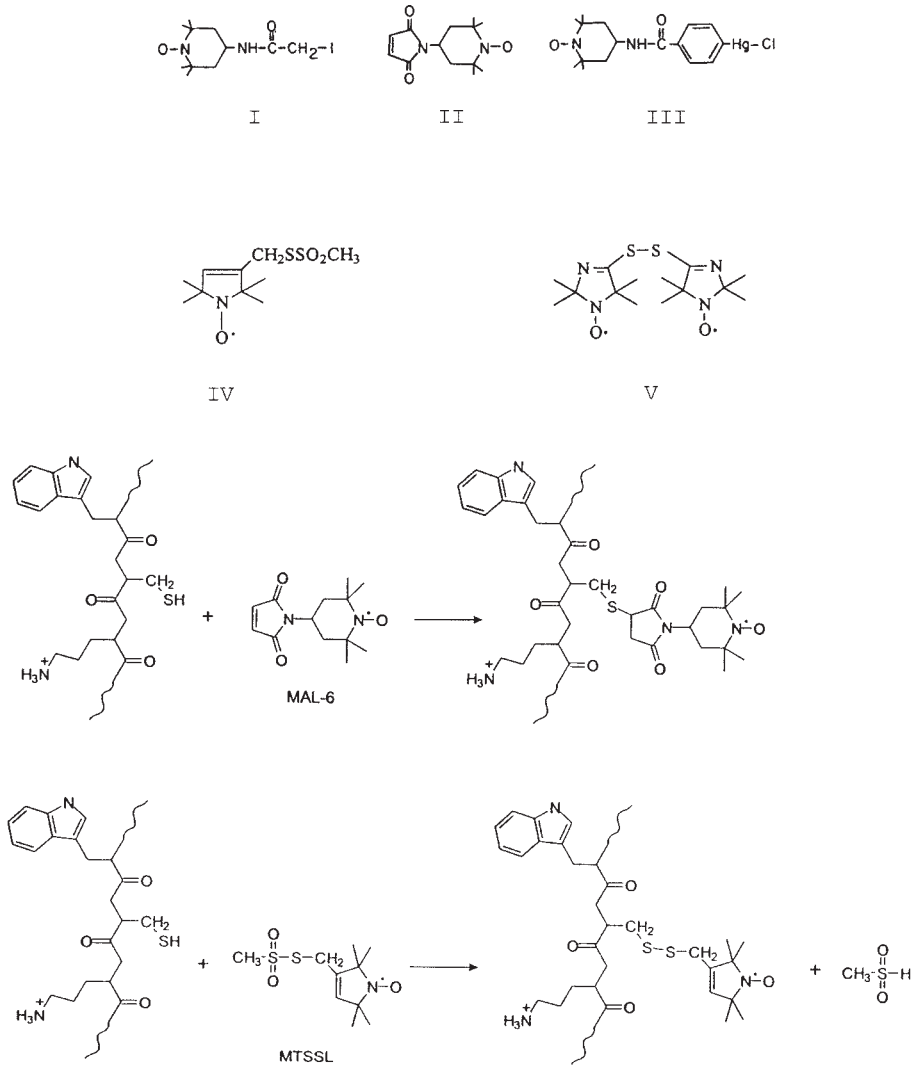


Fig. 2. Representative spin labels for protein labeling. The labels fall into two general classes. Thiol reactive alkylating agents (I and II) and thiol specific labels. Labels I (iodoacetamide spin label) and II (maleimide spin label, sometimes abbreviated MAL–6) react rapidly with –SH groups, but iodoacetamide can also partially alkylate the imidazole nitrogen of histidine, the thioether linkage in methionine and α- and ε-amino groups, depending on the pH. Likewise maleimides can alkylate α-amino groups. Labels III, IV(MTSSL), and V are thiol specific, the latter two resulting in covalent disulfide linkages with the protein. The lower scheme depicts the chemistry of labeling a cysteine on a polypeptide with MAL-6 and MTSSL, respectively. Adapted in part from **ref. *9***.

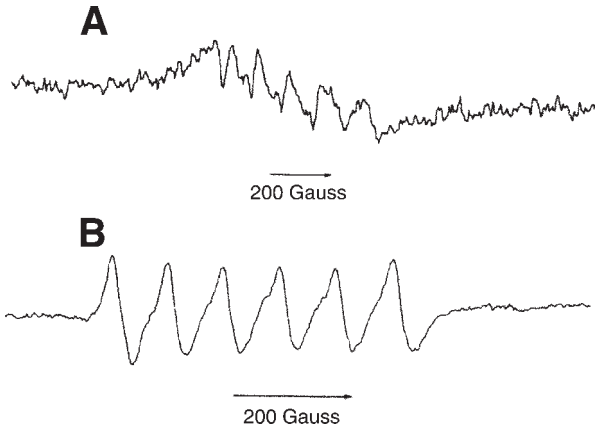


Fig. 3. ESR spectra of Mn(II)-α-lactalbumin complexes at 9 and 35 GHz, pH 7.4, and 0.02 *M* Tris-HCl. **(A)** X-band (9 GHz) ESR spectrum of 0.59 m*M* Mn(II) and 1.2 m*M* bovine α-lactalbumin, 77 K. The spectrum was computer corrected for cavity background and unbound Mn(II) by subtracting a Mn(II) standard in frozen Sephadex G-25 buffer under the same conditions. Conditions were as follows: magnetic field, 3200 G; sweep width, 2000 G; power, 20 mW; modulation, 10 G; sweep time, 2 min; response time, 0.064 s. **(B)** Q-band (35 GHz) ESR spectrum of 0.50 m*M* Mn(II)-4.0 m*M* bovine α-lactalbumin at 10°C. The contribution due to free Mn(II) was 0.8% of the total spectral intensity (Murakami et al., 1982). The spectrum did not narrow when the temperature was increased to ambient temperature. Conditions were as follows: magnetic field, 12,450 G; sweep width, 1000 G; power attenuation, 2 dB; modulation, 6.3 G; response time, 1 s. Reproduced with permission from **ref. *11***.

### *3.3. Data Collection*

1. Scan the sample utilizing an appropriate filter time-constant (typically lengths of seconds to one second) to allow a reasonably noise free scan over the chosen scan range without distorting the spectral lines by sweeping faster than the filter can collect and integrate the signal voltage ***(10)***.
2. Manganese[II]. **Figure 3** shows X-band and Q-band EPR spectra, respectively for 92% bound Mn[II] bound at the strong calcium-binding site of α-lactalbumin, respectively. The X-band spectra (*see* **Fig. 3A**) were determined at 77 K ***(11)***. The spectrum was computer corrected for the 8% unbound Mn[II]. The broad six line spectrum was quite similar to that of aquo-manganese (II) in frozen solution except for the somewhat more pronounced low-field shoulder. This is reflective of slightly distorted, octahedral coordinated cation spectrum. Whereas an octahedral coordinated cation is consistent with known X-ray structures of EF-hand proteins (e.g., parvalbumin) the similarities found with α-lactalbumin are somewhat misleading, given the X-ray structure noted earlier. Whereas the coordination is octahedral from a low-resolution point of view, despite the unique

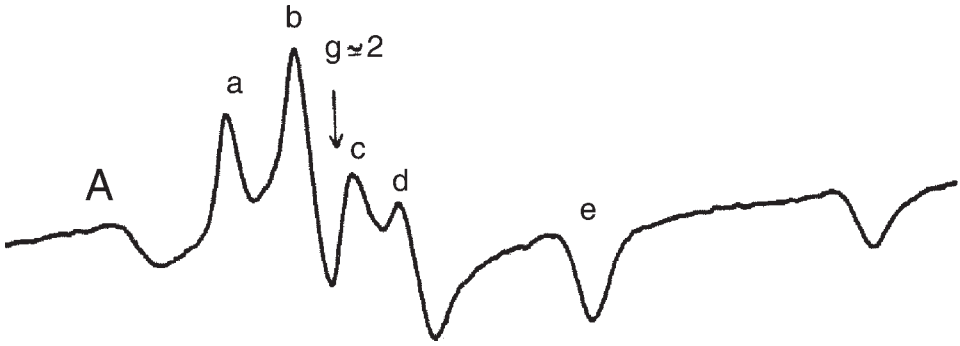


Fig. 4. Q-band EPR spectra of a 2-m*M* Gd(III)-α-LA complex, pH 8.5 (50 m*M* Tris-HCl buffer, $T$ = 273 K). Experimental conditions were as follows: frequency, 34.56; GHz; microwave attenuation, –5 dB; modulation amplitude, 8 G; time constant, 1 s; scan time, 4 min; field set, 12,700 G; scan range, 2500 G. Reproduced with permission from **ref. *11***.

sensitivity of the EPR spectrum to subtleties in electronic structure around manganese, the EPR unfortunately fails here. This was also the problem at 35 GHz (*see* **Fig. 3B**, 10°C), which was almost completely devoid of inhomogeneous broadening contributions from second-order effects because of zero-field splittings ***(12)***. On the other hand, the absence of spectral narrowing with increasing temperature verified that the apparently homogenous line shape was not a result of free, unbound Mn[II] ***(13)***.

3. Gadolinium [III]. Gd[III] is an almost perfect substitute for calcium. Here, the EPR spectrum is interpretable to a greater level of detail than experienced with Mn[II]. **Figure 4** compares Gd[III] α-lactalbumin at Q-band (273 K). The differences and deviations from highly symmetric environments are more discernible particularly of lower temperatures (**Fig. 5**). The most prominent features were clustered near $g$ = 2 constitute a pattern expected for the central ($M_S$ = –1/2) fine structure transition in a crystal field with intermediate rhombic symmetry ***(14)***. The other two features at 11.74 and 13.70 kG (note 1.0T = 10 kG) apparently belong to the outer (satellite) fine transitions. One can determine various energies, quadratic zero-field splitting interactions, and its anisotropy.
4. Vanadyl [IV]. Chasteen showed many years ago that $VO^{2+}$[IV] can substitute for calcium, despite its unusually chemical structure and size ***(15)***. Spectrum 5 shows the X-band spectrum of $VO^{2+}$[IV] α-lactalbumin at 77 K. Analysis of the spectra, by comparing model compounds, indicates that the $VO^{2+}$[IV] was most closely associated with four equatorial oxygen ligands ***(14)***. The linewidth of the $m$ = 1/2 perpendicular line in deuterated water was reduced by 1.7 G, which correlates with a single water ligand in the vanadyl protein complex. This is to be compared with the X-ray crystal structure of human lactalbumin showing two water molecules to a seven-oxygen coordinated calcium ***(16)***.

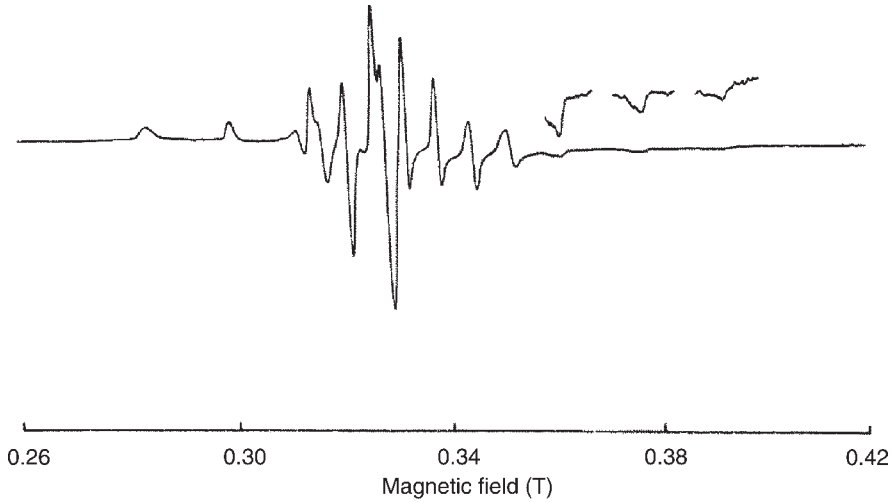


Fig. 5. Low-temperature X-band EPR spectra of a 1 m*M* $VO^{2+}$-α-LA complex, pH 7.4 (10 m*M* HEPES buffer, T = 77 K). Experimental conditions were as follows: frequency, 9.129 GHz; microwave power, 20 mW; modulation amplitude, 10 G; time constant, 0. 128 s; scan time, 4 min; field set, 3400 G; scan range, 1600 G. Inset: Upfield portion of the spectrum at fivefold higher gain. Reproduced with permission from **ref. *13***.

## 4. Notes

1. When studying metal–protein complexes, it is important to try multifrequency experiments. Although X-band (9.5 GHz) spectrometers are very common, the other frequencies (S-band, Q-band, W-band) are available at some of the EPR centers in the United States and various specialized labs around the world.
2. In order to verify that binding occurred at the principal binding site, the EPR spectra must be measured in excess calcium in order to displace the paramagnetic ion. One must still be cautious to check that secondary-site binding occurs between the displaced paramagnetic ion and the protein ***(14)***.
3. More detailed spectral features of the electronic coordination to the cation are more discernible in frozen solution. The most ideal situation would be single crystals of metal-bound protein.
4. Spin-labeled proteins must be exhaustively dialyzed or chromatographed to remove all remnants of free unreacted label. A second, more difficult to reconcile problem is that where the protein is partially proteolyzed or denatured after labeling. Occasionally, gel permeation or sophisticated HPLC methods might separate away the approx 1–5% damaged protein which nonetheless contributes an intense narrow three-line component to the spin-label spectrum.
5. In order to emphasize the covalently bound components of spin-labeled proteins, it is often desirable to slow the motion of the overall protein or local (labeled)

domain. By adding saturated sucrose to the sample, the splittings in the EPR spectra will move outward (i.e., increase), frequently revealing other motional components that might exist ***(3)***. Another technique involves increasing the modulation and receiver gain four- to fivefold, consequently overmodulating the narrow line components and emphasizing the broad slow motional components.

## Acknowledgment

This work was supported in part by grants from the U.S.P.H.S.

## References

1. Berliner, L. J., ed. (1976) *Spin Labeling: Theory and Applications*. Academic, New York, New York.
2. Berliner, L. J., ed. (1979) *Spin Labeling II: Theory and Applications*. Academic, New York, New York.
3. Berliner, L. J. (1980) Using the spin label method in enzymology, in *Spectroscopy in Biochemistry,* vol. II, CRC, West Palm Beach, Florida, pp. 1–56 .
4. Berliner, L. J. (1983) The spin labeling approach to labeling sulfhydryl groups in membrane proteins. *Ann. NY Acad. Sci.* 414 153–161.
5. Eaton, G. R., Eaton, S. S., and Berliner, L. J., eds. (2000) Distance measurements in biological systems by EPR, in *Biological Magnetic Resonance*, vol. 19, Kluwer Academic/Plenum, New York.
6. Musci, G., Koga, K., and Berliner, L. J. (1988) Met-90 spin-labeled bovine α-lactalbumin: ESR and NMR distance measurements. *Biochemistry* **27,** 1260–1265.
7. Berliner, L. J. (1977) Spin labeling in enzymology — spin labeled enzymes and proteins. *Methods Enzymol.* **49G,** 418–480.
8. Koga, K. and Berliner, L. J. (1985) Structural elucidation of a hydrophobic box in bovine a-lactalbumin by NMR: nuclear Overhauser effects. *Biochemistry* **24,** 7257–7262.
9. Feix, J. B. and Klug, C. S. (1998) Site-directed spin labeling of membrane proteins and peptide-membrane interactions, in *Spin Labeling: The Next Millennium, Biological Magnetic Resonance*, vol. 14 (Berliner, L. J., ed.), Plenum, New York, pp. 251–281.
10. Jost, P. C. and Griffith, O. H. (1976) Instrumental aspects or spin labeling, in *Spin Labeling: Theory and Applications* (Berliner, L. J., ed.), Academic, New York.
11. Berliner, L. J., Ellis, P. D., and Murakami, K. (1983) Mn(II) and 113-Cd NMR evidence for the nature of the Ca(II) binding site in α-lactalbumin. *Biochemistry* **22,** 5061–5063.
12. Buttlaire, D. H., Reed, G. H., and Himes, R. (1975) Electron paramagnetic resonance and water proton relaxation rate studies of formyltetrahydrofolate synthetase-manganous ion complexes. Evidence for involvement of substrates in the promotion of a catalytically competent active site. *J. Biol. Chem.* **250,** 261–270.
13. Reed, G. H. and Markham, G. D. (1984) EPR of Mn(II) complexes with and enzymes other proteins, in *Biological Magnetic Resonance*, vol. 6 (Berliner, L. J., ed.), Plenum, New York, pp. 73–142.

14. Musci, G., Reed, G. H., and Berliner, L. J. (1986) An electron paramagnetic resonance study of metal ion binding to bovine α-lactalbumin. *J. Inorg. Biochem.* **26,** 229–236.
15. Chasteen, N. D. (1995) Vanadium-protein interactions. *Metal Ions Biol. Syst.* **31,** 231–247.
16. Acharya, K. R., Ren, J., Stuart, D. I., and Phillips, D. C. (1991) Crystal structure of human α-lactalbumin at 1. 7 Å resolution. *J. Mol. Biol.* **221,** 571–581.

# 17

# Cadmium-113 and Lead-207 NMR Spectroscopic Studies of Calcium-Binding Proteins

**Teresa E. Clarke and Hans J. Vogel**

## 1. Introduction

Ideally, direct NMR spectroscopic studies of the structure and dynamics of a metal binding site would utilize the ion that is naturally bound to the protein. For calcium-binding proteins, the NMR active isotope $^{43}Ca$ is a quadrupolar (I = 7/2) nucleus that produces broad peaks when bound to a protein ***(1–4)***. Although much useful information can be gleaned from studies of this nucleus ***(3,4)***, it is impossible to resolve the signals for individual calcium-binding sites in a protein with multiple-binding sites. However, isomorphous replacement of the calcium ion with an ion with more favorable NMR properties often allows the resolution of NMR signals for individual sites and provides valuable insight into the coordination of the ion within the binding site from the observed shifts. $^{113}Cd$, and to a lesser extent, $^{207}Pb$, are two such I = 1/2 metal nuclei that have been successfully used to characterize the $Ca^{2+}$-binding properties of a variety of metalloproteins ***(5–8)***.

Although $Cd^{2+}$ and $Pb^{2+}$ are toxic metal ions and not of direct biological relevance, they can effectively substitute for $Ca^{2+}$ in metalloproteins, allowing them to retain similar structures and function. $Cd^{2+}$ has a filled *d*-shell orbital and can form complexes with a variety of conformations and number of ligands. Relative to $Ca^{2+}$ (r = 0.99 Å), $Cd^{2+}$ has a very similar ionic radius (r = 0.97 Å), whereas $Pb^{2+}$ is slightly larger (r = 1.20 Å).

For both $^{113}Cd$ and $^{207}Pb$, the chemical shift of the NMR signal is highly dependent on the ligand environment of the metal ion. An increase in the polarizability of the ligand results in a decrease in nuclear shielding (i.e., $S < N < O$), such that oxygen-coordinated metal signals resonate the furthest upfield. Typi-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel © Humana Press Inc., Totowa, NJ

cal calcium-binding sites in proteins are usually made up of oxygen ligands, so their NMR signals are found in the upfield portion of the spectra. The large chemical shift window available in both $^{113}Cd$ and $^{207}Pb$ NMR can also distinguish subtle differences between motifs in $Ca^{2+}$-binding sites ***(3–8)***.

This chapter will focus on simple one-dimensional (1D) $^{113}Cd$ and $^{207}Pb$ NMR spectroscopic techniques in order to demonstrate the ease of distinguishing between different calcium-binding motifs. Other NMR studies involving these nuclei, including solid-state NMR studies or 2D heteronuclear NMR spectroscopy, can in principle, also be useful to characterize $Ca^{2+}$-binding proteins, but are beyond the scope of this work. Cadmium has been substituted for a wide range of metal ions other than $Ca^{2+}$, and the reader is referred to a number of comprehensive reviews on the use of cadmium-113 NMR spectroscopy for studies of such metalloproteins ***(5–7)***.

## 2. Materials

1. Acid-washed glassware and plasticware should be prepared by washing extensively in 1 *M* HCl and rinsing thoroughly in distilled $H_2O$.
2. For calcium removal from the protein, one can use a 10 DG column (Bio-Rad Laboratories) and 50 m*M* ammonium bicarbonate and 0.5 *M* ethylenediaminetetracetic acid (EDTA), pH 8.0. Alternately, the protein can be passed through a Chelex column, eluting with 50 m*M* ammonium bicarbonate (*see* **Note 1**).
3. A widebore Bruker AM400 instrument equipped with a 10-mm broadband probe with variable temperature control capabilities (*see* **Note 2**). Higher or lower field instruments can be used with appropriate changes to the parameters (*see* **Note 3**).
4. A 100-m*M* stock solution of $^{113}Cd(ClO_4)_2$ is prepared by dissolving the $^{113}CdO$ (95%) (Cambridge Isotope Laboratories) in concentrated HCl, heating to dryness under a stream of nitrogen with low heat, and dissolving the residue in the appropriate volume of 200 m*M* $HClO_4$ in $D_2O$ (99.9%) (Cambridge Isotope Laboratories) ***(9)***. Alternatively, a 100-m*M* solution of $^{113}CdSO_4$ can be prepared as the stock solution (*see* **Note 4**) by dissolving in 3 *M* $H_2SO_4$, heating to dryness as above, and taking the powder up in a correct volume of 50/50 $D_2O/H_2O$ and neutralizing with NaOH to pH 6.0 ***(10)***.
5. A 100-m*M* stock solution of $^{207}Pb(NO_3)_2$ (91.6%) (Oak Ridge National Laboratories) is made up in $D_2O$ to be used as a titrant and chemical shift reference ***(8)***.
6. Materials for limited proteolysis of the $Ca^{2+}$-binding proteins, as described in Chapter 15 in this volume ***(11)***.

## 3. Methods

### 3.1. Sample Preparation

The appropriate amount of a purified, lyophilized calcium-binding protein is dissolved in 1 mL 50 m*M* ammonium bicarbonate and 150 μL 0.5 *M* EDTA, pH 8.0, and desalted by two passes through a 10 DG column, then freeze dried.

A 2-mL protein sample is prepared in 90% $H_2O$, 10% $D_2O$, 100 m*M* KCl, and the pH is adjusted to 7.5 (or desired pH) with small amounts of 0.5 *M* KOD or 0.5 *M* DCl (*see* **Note 5**). Ideally, the concentration of the protein sample (determined spectrophotometrically) should be at least 1 m*M* for obtaining a reasonable signal-to-noise ratio.

### 3.2. Spectral Acquisition

Typical $^{113}$Cd NMR spectra (88.75 MHz) are acquired locked with a pulse length of 45–60°, sweep widths of 30 kHz, repetition rates of 0.50–0.60 s, and 4–8 k data points. When acquired in this fashion, the spectra are not usually fully relaxed and quantification by integration of the area under the peak is not reliable. In order to be able to make a relative comparison of the amount of $Cd^{2+}$ under each peak, the $^{113}$Cd nucleus must be fully relaxed; to determine an appropriate delay time between pulses, the delay time is increased until the amplitude of the signal ceases to increase. Each free induction decay (FID) is generally zero filled once and processed with an exponential multiplication function typically resulting in a line broadening of 30 Hz. The external chemical shift reference is either 100 m*M* $^{113}Cd(ClO_4)_2$ or $^{113}CdSO_4$ in $D_2O$.

$^{207}$Pb NMR (83.45 MHz) acquisition parameters include 60–80° pulses, sweep widths of 100 kHz, 0.40–0.55 s between pulses, and 16 k data points. Large sweep widths are employed because of the wide chemical shift window of $^{207}$Pb, but unequal excitation of the signal could result ***(8)***. During processing, spectra are usually left-shifted, giving an effective dead time of 50–70 μs, zero filled once, with an exponential line broadening of 300–500 Hz. $^{207}Pb(NO_3)_2$ in $D_2O$ can be used as the external chemical shift reference.

The sample is inserted into the instrument and a number of scans are taken. The number of scans required depends upon the concentration of the sample, the desired noise level of the spectra, and the time allotted for the experiment. For $^{113}$Cd NMR, this is typically 50,000 scans for a 1-m*M* protein solution, whereas 100,000 scans are required for $^{207}$Pb NMR spectra at the same concentration, taking a full day for data acquisition. During a metal-ion titration, the pH of the sample is checked after each addition of metal solution before the spectra are obtained. **Figure 1** shows a $^{113}$Cd NMR spectrum obtained for calbindin $D_{9K}$ (ICaBP, intestinal calcium-binding protein), and, for comparison, typical chemical shifts for several other calcium-binding proteins including calmodulin (CaM), troponin C (TnC), parvalbumin, and lactalbumin ***(12)***. **Figure 2** shows a $^{207}$Pb NMR spectrum of CaM. Two narrow peaks present in the spectrum, correspond to slow exchange binding with two calcium-binding sites (III and IV, assigned by proteolytic fragments, *see* **Subheading 3.4.**), whereas the other broad resonances are attributed to the other two binding sites (I and II) of CaM ***(8)***. During titration experiments with CaM, sites III and IV

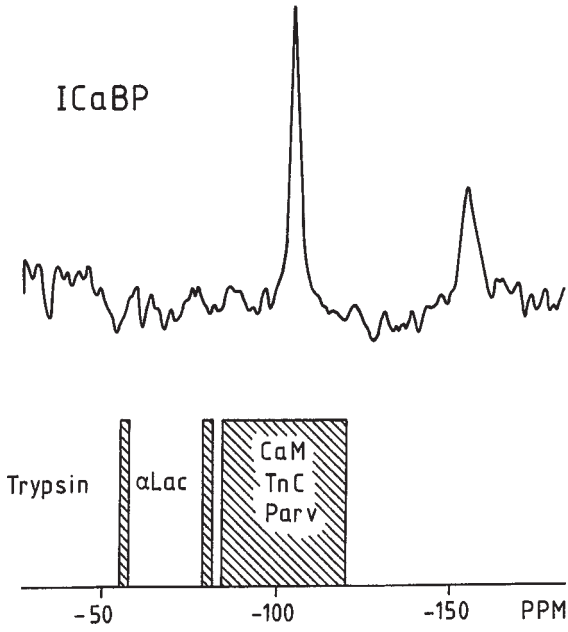


Fig. 1. $^{113}Cd$ NMR spectrum of calbindin (ICaBP) and typical chemical shifts observed for other calcium-binding proteins including calmodulin, troponin C, parvalbumin, and lactalbumin. Because these calcium-binding proteins only use oxygen to coordinate metal ions, they are found in a similar range ***(12)***.

are simultaneously filled first, before sites I and II become occupied. Simultaneous filling of two sites in one lobe of a protein often can indicate positive cooperativity ***(2,3)***. In addition, this experiment shows that the C lobe of CaM has a higher affinity for $Cd^{2+}$ than the N lobe.

### 3.3. Chemical Exchange of Ions

Resonances for $^{113}Cd^{2+}$ (or $^{207}Pb^{2+}$) ion-binding sites may be absent from the spectra at room temperature. Because the chemical shift range is so large, chemical exchange can rapidly broaden a peak beyond detection. For CaM and TnC, only the two carboxy-terminal sites are detected while the signals from the ions bound to the amino-terminal half are too broad to be seen at room temperature. To change the rate of chemical exchange, spectra can be acquired at varying temperatures for both $^{113}Cd$ and $^{207}Pb$. Acquisition parameters for low-temperature experiments are similar but the chemical shift of the external standard should be reevaluated at each temperature. In order to study the effect of exchange on the spectra, it is often useful to record spectra at a number of temperatures between 5°C and 50°C. For example, at lower temperatures, two

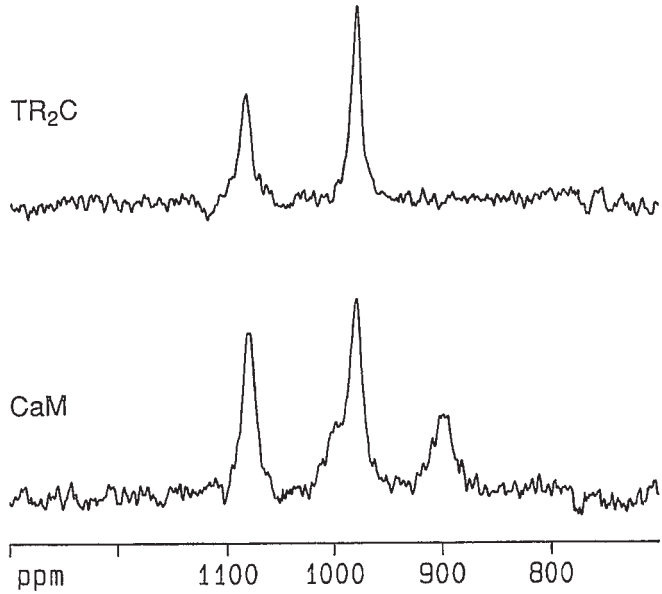


Fig. 2. $^{207}$Pb NMR spectra of intact CaM and its *C*-terminal domain fragment, $TR_2C$ (residues 78–148). Four signals are observed in the spectra; the two sharp signals can be assigned to the C lobe by lining up with the two peaks for the half-molecule of CaM, whereas the two broad peaks can be attributed to the other two calcium-binding sites in the N lobe ***(8)***.

additional broad peaks not present in the room temperature spectra of TnC, which correspond to the *N*-terminal domain-binding sites, can be observed.

However, even at low temperatures, additional resonances for the two weak binding sites of CaM are not present in the spectra (*see bottom panel* of **Fig. 3**). In this case, it is likely that the C-domain of the protein is static, whereas the N-domain is rapidly changing conformation. If the *N*-terminal $^{113}Cd^{2+}$ ions have different shifts in these different conformations, the spectra obtained at different exchange rates can be simulated (*see* **Fig. 4**). The spectra obtained for CaM at room temperature match the spectrum calculated for an exchange rate ($k_{exch}$) of $10^4$ $s^{-1}$ ***(2,3)***. Often, it is thought that the exchange process responsible for the broadening involves free $Cd^{2+}$ in solution; however, in the case of CaM, when a fifth equivalent is added, a free $Cd^{2+}$ signal can be observed, which is in slow exchange (data not shown). Hence, a conformational exchange process must be occurring in the *N*-terminal domain, as indicated in **Fig. 4**. When a peptide binds to CaM, the N-domain adopts one unique conformation and with the exchange process eliminated, four peaks appear in the spectra for the four calcium-binding sites of CaM (*see top panel* of **Fig. 3**) ***(13)***.

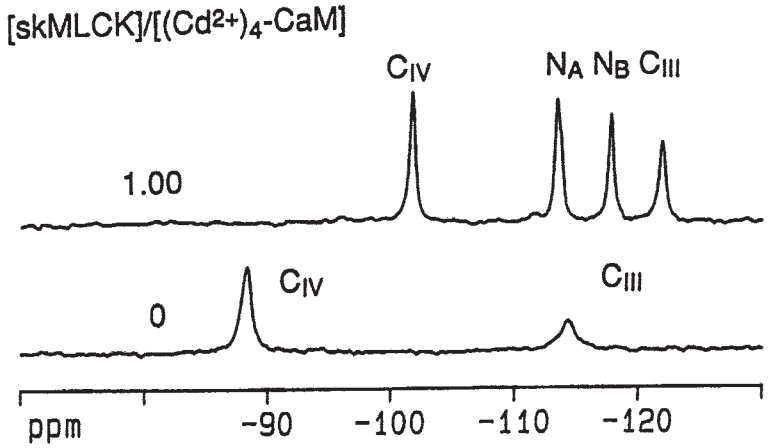


Fig. 3. $^{113}$Cd NMR spectra of CaM and a 1:1 complex with MLCK (myosin light-chain kinase) peptide. Only two signals are present for cadmium-saturated CaM, corresponding to the two *C*-terminal sites, whereas the two signals for the *N*-terminal sites are broadened beyond detection. Upon addition of peptide, four signals at different chemical shifts appear, matching each of the four sites. Binding of the peptide abolishes the conformational exchange process in the *N*-terminal domain of CaM (*see* **Fig. 4**) ***(13)***.

### *3.4. Using Proteolytic Fragments to Assign Peaks to Binding Sites*

Individual domains from many calcium-binding proteins can be isolated by limited proteolytic degradation with various enzymes (*see* **Note 6**) ***(11)***. Structurally intact binding sites produce spectra with very similar chemical shifts as the intact sites in the whole protein ***(13,14)***. Ideally, each metal-ion binding site of the protein would be isolated for efficient identification of each resonance. However, it is sometimes possible to assign resonances with fragments containing more than one binding site. For example, calmodulin can be digested with either thrombin or trypsin to produce different fragments containing varying numbers of binding sites ***(15)***. NMR samples containing the proteolytic fragments of the protein are prepared similarly to those for the intact protein and experimental parameters also remain the same.

An example is shown in the $^{207}$Pb NMR spectra in **Fig. 2**, where the two narrow signals for CaM clearly arise from the C lobe ($TR_2C$, residues 78–142) of the protein, and by elimination, the other two signals are from the N-lobe of the protein ($TR_1C$, residues 1–77). Slight differences are sometimes found between the chemical shifts for the fragments and whole protein. These may be a result of slight perturbations in the $Ca^{2+}$-binding loops of the binding site, caused by proteolytic cleavage.

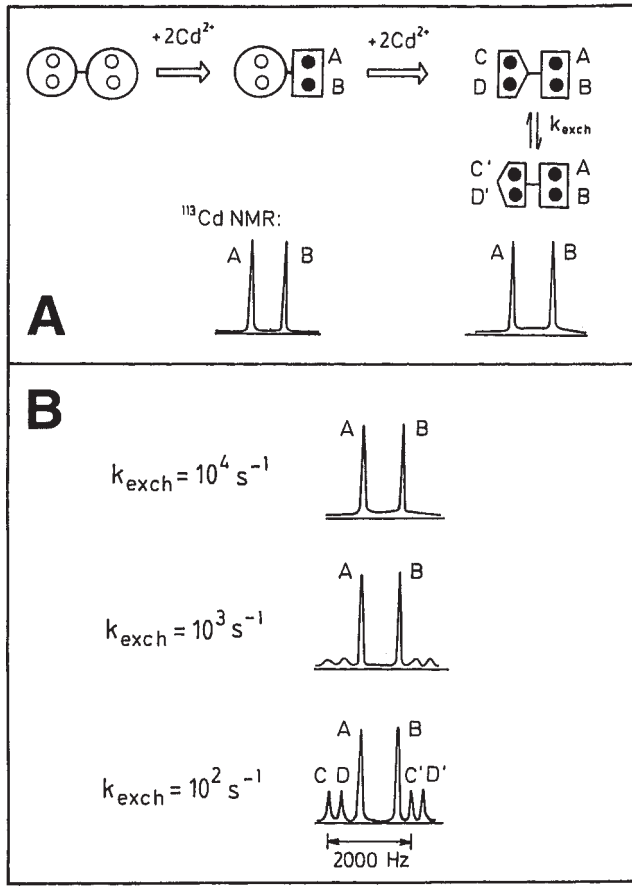


Fig. 4. Simulation of $^{113}Cd$ NMR spectra with different exchange rates. **(A)** Schematic of the different situations that may exist for the protein if the N-domain could change conformation. **(B)** The simulated spectra at different exchange rates if the N-domain $^{113}Cd^{2+}$ ions have different chemical shifts (30–35 ppm) ***(2,3)***.

### 3.5. Other Assignment Strategies

Other methods for assignment of the signals to each binding site in the protein exist. Differences in the relative affinity of each site for various metal ions and other factors that perturb one site over the other can be exploited. For example, in **Fig. 5**, the signals in the $^{113}Cd$ NMR spectra of parvalbumin were assigned when $Lu^{3+}$ displaced the $^{113}Cd^{2+}$ from the EF site, but not the CD site ***(16)***. Labeling of specific sites within the binding pocket for the metal ion often alters the corresponding signal, allowing identification. Fluorescent derivatives of skeletal troponin C were utilized to confirm assignment of $^{113}Cd$

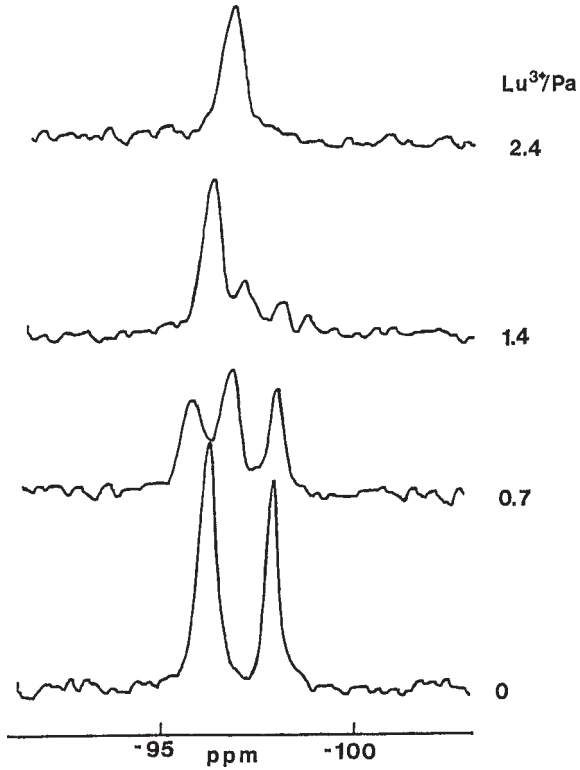


Fig. 5. $^{113}Cd$ NMR spectra of carp parvalbumin titrated with the diamagnetic lanthanide $Lu^{3+}$. $Lu^{3+}$ displaces $Cd^{2+}$ from the EF site of parvalbumin but not the CD site. The chemical shift of the peak for the CD site depends on the occupancy of the EF-hand with either $Cd^{2+}$ or $Lu^{3+}$ (slow exchange). This resonance shifts slightly because of interactions between the free $Cd^{2+}$ and a third weak metal-ion binding site on parvalbumin (fast exchange) ***(16)***.

resonances by Ellis et al. ***(17)***. Sometimes paramagnetic metal ions or nitroxide probes can be used to aid in making assignments. Finally, important liganding residues in a binding site could also be mutated to prevent metal-ion binding, causing the disappearance of the corresponding signal.

### 3.6. Special Considerations

When using substitute metal ions such as $Cd^{2+}$ or $Pb^{2+}$ in lieu of the native ion $Ca^{2+}$, there is always a concern that the protein does not undergo the same conformational changes upon metal-ion binding. This can be studied by proton NMR titration experiments, for $Cd^{2+}$ and $Ca^{2+}$, one usually obtains the same results ***(3)***, whereas $Pb^{2+}$ is often somewhat different ***(8)***. Also, highly resolved

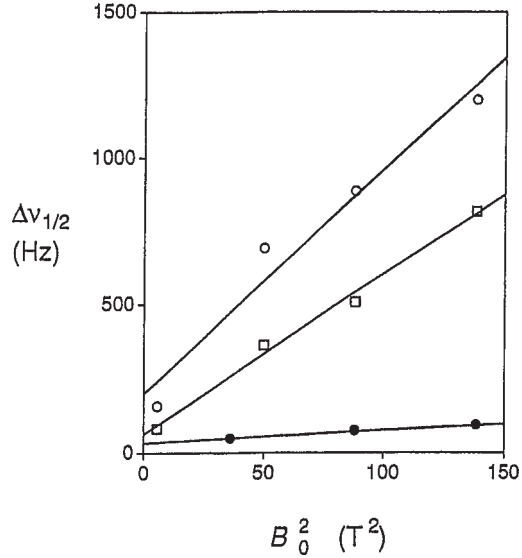


Fig. 6. Relationship of the CaM-bound $^{113}Cd$ and $^{207}Pb$ linewidths ($\Delta\nu_{1/2}$) with the square of the magnetic field strength ($B_o$). ○ and □ from $^{207}Pb$ CaM peaks at δ = 983 ppm and δ = 965 ppm, respectively. ● from $^{113}Cd$ CaM signal at δ = –114.5 ppm.

crystal structures of some calcium-binding proteins with $Cd^{2+}$ are identical to those with $Ca^{2+}$, lending credence to the substitution method ***(18)***.

## 4. Notes

1. It is not always trivial to completely remove EDTA or EGTA once they have been introduced into the protein preparation so treatment with Chelex is preferred whenever possible. Peaks from remaining EDTA or EGTA contamination can sometimes be detected in the NMR spectra.
2. For lower-affinity $Ca^{2+}$-binding sites, chemical exchange rates between free and bound $Cd^{2+}$ may cause signals to be too broad for detection. On the other hand, broad resonances could result from conversion between two or more protein conformations (*see* **Fig. 4**). These two possibilities must be considered; temperature variation experiments are very useful in this respect. The chemical exchange rate for $^{207}Pb^{2+}$ bound to calmodulin is also very temperature sensitive, with peaks disappearing above or below ambient temperature ***(8)***.
3. NMR sensitive nuclei with a large electron cloud such as $^{113}Cd$ or $^{207}Pb$ often relax through a mechanism known as chemical shift anisotrophy (CSA). For protein-bound nuclei, this effect becomes rather substantial and should be considered when selecting an NMR spectrometer, which will give optimal results**.** **Figure 6** shows the consequence of CSA on $^{113}Cd$ and $^{207}Pb$ NMR spectra of CaM, recorded on 100 MHz, 300 MHz, 400 MHz, and 500 MHz instruments.

The linewidth increases with the square of the magnetic field strength, hence, better spectra with narrower linewidths can often be obtained at somewhat lower fields. The effect is less dramatic for $^{113}Cd$ than for $^{207}Pb$ (*see* **Fig. 6**). Measurements of relaxation times are required to properly analyze these relaxation phenomena ***(8,9)***.

4. In recent years, the use of perchloric acid has been discouraged in certain laboratories. Either $^{113}CdSO_4$ or $^{113}Cd(ClO_4)_2$ can be successfully used as the external standard and titrant, but slight differences can be obtained in the chemical shift with various counter ions.
5. Buffers are not normally used in sample preparation for $^{113}Cd$ NMR spectroscopy since $Cd^{2+}$ ions will interact with ions such as Tris and $Cl^-$, causing changes in linewidth and chemical shift of the resonances. It should also be noted that all NH groups on the protein will weakly interact with $Cd^{2+}$, causing the free $Cd^{2+}$ signal to broaden. Such very weak binding sites will contribute to the exchange processes, making it difficult to obtain reasonable $K_d$'s for a weak $Cd^{2+}$ binding site in a protein from $^{113}Cd$ NMR spectroscopy.
6. In order to properly assign the resonances, domain fragments of the protein must retain the characteriztics of the intact protein. The structure of the fragment should be similar, with the same conformational changes induced when the metal-ion binds. Often, the structure and associated metal ion induced changes of the proteolytic fragments are examined by a number of experiments to confirm their identity to each domain of the intact protein. These include $^1H$ NMR spectroscopy, circular dichroism experiments or Fourier transform infrared spectroscopy with titrations of each metal ion.

## Acknowledgments

This work was supported by the Medical Research Council (MRC) of Canada and the Alberta Heritage Foundation for Medical Research (AHFMR). T. E. Clarke is supported by a Doctoral Research Award from MRC Canada and H. J. Vogel is an AHFMR Scientist. We thank Dr. T. Drakenberg and Dr. J. Aramini for collaborations and numerous insightful discussions.

## References

1. Vogel, H. J., Drakenberg, T., and Forsén, S. (1983) Calcium binding proteins, in *NMR of Newly Accessible Nuclei* (Laszlo, P., ed.), Academic, New York, pp. 157–192.
2. Forsén, S., Vogel, H. J., and Drakenberg, T. (1986) Biophysical studies of calmodulin, in *Calcium and Cell Function*, vol. VI (Cheung, W. Y., ed.), Academic, New York, pp. 113–157.
3. Vogel, H. J. and Forsén, S. (1987) NMR studies of calcium binding proteins. *Biol. Magn. Res.* **7,** 245–307.
4. Drakenberg, T. (2002) Calcium-43 NMR of calcium-binding proteins, in Calcium-Binding Protein Protocols: Methods and Techniques, Vol. 2 (Vogel, H. J., ed.), Humana Press, Totowa, New Jersey, pp. 217–230.

5. Summers, M. F. (1988) Cadmium-113 NMR spectroscopy of coordination compounds and proteins. *Coord. Chem. Rev.* **86,** 43–134.
6. Coleman, J. E. (1993) Cadmium-113 nuclear magnetic resonance applied to metalloproteins. *Methods Enzymol.* **227,** 16–43.
7. Öz, G., Pountney, D. L., and Armitage, I. M. (1998) NMR spectroscopic studies of I = 1/2 metal ions in biological systems. *Biochem. Cell Biol.* **76,** 223–234.
8. Aramini, J. M., Hiraoki, T., Yakawa, M., Yuan, T., Zhang, M., and Vogel, H. J. (1996) Lead-207 NMR: a novel probe for the study of calcium-binding proteins. *J. Biol. Inorg. Chem.* **1,** 39–48.
9. Aramini, J. M., Hiraoki, T., Ke, Y., Nitta, K., and Vogel, H. J. (1995) Cadmium-113 NMR studies of bovine and human -lactalbumin and equine lysozyme. *J. Biochem.* **117,** 623–628.
10. Sudmeier, J. L., Bell, S. J., Storm, M. C., and Dunn, M. F. (1981) Cadmium-113 nuclear magnetic resonance studies of bovine insulin: two-zinc insulin hexamer specifically binds calcium. *Science* **212,** 560–562.
11. Brokx, R. D. and Vogel, H. J. (2002) Proteolytic fragments of calcium binding proteins. *Methods Mol. Biol.,* pp. 183–184.
12. Vogel, H. J., Drakenberg, T., Forsén, S., O'Neill, J. D., and Hofmann, T. (1985) Structural differences in the two calcium binding sites of porcine intestinal calcium binding protein. *Biochemistry* **24,** 3870–3876.
13. Zhang, M., Yuan, T., Aramini, J., and Vogel, H. J. (1995) Interaction of calmodulin with its binding domain of rat cerebellar nitric oxide synthase: a multinuclear NMR study. *J. Biol. Chem.* **270,** 20,901–20,907.
14. Andersson, A., Forsén, S., Thulin, E., and Vogel, H. J. (1983) Cadmium-113 nuclear magnetic resonance studies of proteolytic fragments of calmodulin: assignment of strong and weak cation binding sites. *Biochemistry* **22,** 2309–2313.
15. Drakenberg, T., Forsén, S., Thulin, E., and Vogel, H. J. (1987) The binding of $Ca^{2+}$, $Mg^{2+}$, and $Cd^{2+}$ to tryptic fragments of skeletal muscle troponin C: Cadmium-113 and proton NMR studies. *J. Biol. Chem.* **262,** 672–678.
16. Drakenberg, T., Swärd, M., CavJ, A., and Parello, J. (1985) Metal-ion binding to parvalbumin: A $^{113}Cd$ NMR study of the binding of different lanthanide ions. *Biochem. J.* **227,** 711–717.
17. Ellis, P. D., Strang, P., and Potter, J. D. (1984) Cadmium-substituted skeletal troponin C: cadmium-113 NMR spectroscopy and metal binding investigations. *J. Biol. Chem.* **259,** 10,348–10,356.
18. Swain, A. L., Kretsinger, R. H., and Amma, E. L. (1989) Restrained least squares refinement of native calcium and cadmium-substituted carp parvalbumin using x-ray crystallographic data at 1.6 Å resolution. *J. Biol. Chem.* **264,** 6620–6628.

# 18

# Calcium-43 NMR of Calcium-Binding Proteins

**Torbjörn Drakenberg**

## 1. Introduction

The $Ca^{2+}$-ion, with its closed-shell electronic structure, has very few spectroscopic properties that are useful in studies of calcium-binding proteins and their interactions with calcium ions. Therefore, it is even more important to exploit the few possibilities for spectroscopy that are available. The most abundant calcium isotope, $^{42}Ca$, has a spin quantum number I = 0 and is therefore not observable by NMR. However, $^{43}Ca$ has I = 7/2 and is in principle NMR active. The natural abundance of this istope is only 0.15% and $^{43}Ca$ NMR can therefore most likely never be used to study biological systems without isotope enrichment. Nowadays, $^{43}Ca$ enriched to 50% is readily available, though at a rather high cost. It seems that most research groups working with calcium-binding proteins consider the $^{43}Ca$ isotope as an exotic one and only a limited number of groups have been involved in $^{43}Ca$ NMR studies of biological systems ***(1–6)***. Only the Swedish group in Lund and the Japanese group in Sendai have been using $^{43}Ca$ NMR more consistently.

Below I will first present, very briefly, what is special about the application of $^{43}Ca$ NMR as compared to typical spin = 1/2 nuclei like $^{1}H$, $^{13}C$, and $^{15}N$ and than discuss some typical studies.

## 2. Experimental and Theoretical Considerations

Because $^{43}Ca$ is a quadrupolar nucleus with a quite high spin quantum number (I = 7/2) it has properties that are in some respects different from the dipolar nuclei that are mostly considered in studies of macromolecules. From an experimental point of view it is essential to recognize that sensitivity is the most important fact to consider because the natural width of the $^{43}Ca$ NMR resonance from a $^{43}Ca^{2+}$ ion bound to a macromolecule is hundreds, if not thou-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

sands, of Hertz broad. In a way that should simplify the life for the experimentalist because he does not have to compromise between sensitivity and resolution. On the other hand, there seem to be no probes for this purpose available from the NMR vendors.

Even if a high sensitivity probe is available, spectra have to be accumulated for a long time, often more than $10^5$ transients. This puts high demands on the "ring-down time" of the probe. The specifications by the vendors provide no guidance on this. They have most likely never had to run a 1D spectrum with one million transients. The simple remedy for this problem of course would be to increase the dead time, i.e., the time between the end of the last pulse and the acquisition of the first data point, until there is no effect from probe ringing. Unfortunately, for $^{43}$Ca NMR, this has most likely also removed the signal you want to observe. A better way is therefore to use the so-called RIDE sequence *(7)* to minimize the effect from probe ringing. This has worked quite well in our hands and made the observation of resonances that are approx 1000 Hz broad possible even for as much as $10^6$ transients.

The basic theory for NMR may be found in standard text books for NMR (*see*, e.g., Ernst, Bodehausen, and Wokaun *[8]*) and will not be presented here. Some aspects particular for quadrulopar nuclei that will be of importance for the following discussion will be dealt with here.

NMR spectra of quadrupolar nuclei like $^{43}$Ca will be dominated by the effects of the quadrupole moment. In this chapter, I will consider only isotropic solutions, and furthermore limit myself to present only the effects on the linewidth (relaxation) and not on the chemical shift, because such effects have not yet been observed for $^{43}$Ca even though they can be foreseen. In isotropic solutions the linewidth of the $^{43}$Ca resonance will be dominated by the quadrulopar relaxation mechanism and is determined by the quadrupole coupling constant, $\chi = e^2qQ$, and the correlation time $\tau_c$. Only when the extreme narrowing condition applies ($\omega_0\tau_c < 1$) will the relaxation be truly mono-exponential. Under this condition the longitudinal, $R_1$, and transverse, $R_2$, relaxation rates will be equal and given by **Eq. 1**.

$$R_1 = R_2 = 3\pi^2/10*(2I + 3)/[I^2(2I - 1)]*\chi^2(1 - \eta^2/3)\tau_c \qquad (1)$$

where $\chi$ is the quadrupole coupling constant, $\eta$ is an asymmetry parameter for the field gradient, and $\tau_c$ is the correlation time. $(1–\eta^2/3)$ is in most cases close to unity and will therefore be neglected. Even for a very small protein, the extreme narrowing condition is not strictly fulfilled unless the $Ca^{2+}$-ion has some extra mobility in its binding site. Therefore, **Eq. 1** will almost never be strictly valid and $R_1 \neq R_2$.

When $\omega_0\tau_c$ is no longer small compared to unity but still less than 1.5 (so-called nearly extreme narrowing) the apparent relaxation rates $R_1$ and $R_2$

are different and field dependent ***(9)***. For systems where the $^{43}Ca^{2+}$ ion undergoes exchange between a large excess of free, solvated ions, and protein bound ions the apparent relaxation rates will be mono-exponential even though $\omega_0\tau_c$ >1.5 for the protein bound ions. Halle and Wennerström ***(10)*** have derived approximate analytical expressions for the relaxation rates in the near extreme narrowing regime for spin I = 5/2 and I = 7/2 nuclei. The longitudinal and transverse relaxation rates of the major relaxation components are to first-order given by **Eqs. 2** and **3**.

$$R_1 = 3\pi^2/10*\chi^2*(2I + 3)/[I^2(2I - 1)]*[0.2*J_1 + 0.8*J_2] \quad (2)$$

$$R_2 = 3\pi^2/10*\chi^2*(2I + 3)/[I^2(2I - 1)]*[0.3*J_0 + 0.5*J_1 + 0.2*J_2] \quad (3)$$

where the spectral densities are given by

$$J_q = \tau_c/[1 + (q*\omega_0*\tau_c)^2]$$

These equations are analogous to those derived by Bull ***(11)*** for spin I = 3/2 and $\omega_0\tau_c$ <1.5. Halle and Wennerström ***(10)*** have shown that for both I = 5/2 and I = 7/2, **Eqs. 2** and **3** are good approximations for the relaxation rates when they are measured with pulse techniques. When $R_2$ is estimated from the linewidth of the signal, however, considerable errors may be introduced. The nonLorentzian line shape of the NMR signal causes this.

For longer correlation times, it is no longer possible to describe the relaxation as mono-exponential and one has to resort to numerical solutions. This will, however, not be considered in this chapter. *See*, e.g., Drakenberg et al. ***(12)*** for more details and references.

NMR spectra are dependent on dynamical effects in vastly different rate windows. As aforementioned, the relaxation rate and therefore also the linewidth of the resonances depends on the correlation time, in the ns range. The appearance of the NMR spectra will also depend on chemical exchanges in the ms to s range when the observed nucleus takes part in such an exchange. The effect will be small as long as the exchange rate $k_{ex}$ is small compared to the relaxation rate of the observed resonance. Traditionally, chemical exchange rates have been measured for resonances with different chemical shifts ***(13)***, however, they can equally well be measured for resonances with the same chemical shifts, but with widely different relaxation rates. This is a very common situation for quadrupolar ions exchanging between the free, solvated, ions with a relaxation rate in the order of 1 $s^{-1}$ and ions bound to a protein with a relaxation rate in the order of $10^3$ $s^{-1}$. Complete band shape calculations based on the equations derived by McConnell ***(14)*** are very useful. In our laboratory, this has been used on several occasions for $^{43}Ca$, as well as for $^{25}Mg$ NMR ***(15–25)*** and below follows an outline of how relevant exchange parameters can be obtained from the temperature dependence of the NMR spectra.

1. Bloch equations modified to take exchange into account are used to derive the appropriate band shape equation for the actual exchange system. This approach is applicable both for a simple two-site exchange between free ions and ions bound to a single site, as well as for more complex exchange systems with several nonidentical sites on the protein.
2. The exchange rates are assumed to obey the transition state theory and their temperature dependencies are obtained from

$$k_{ex} = (kT/h)\exp[-(\Delta H^{\#} - T^{*}\Delta S^{\#})/RT] \quad (4)$$

   In most cases the data will not be sufficiently reliable to determine $\Delta S^{\#}$ and $\Delta H^{\#}$ simultaneously, in which case $\Delta S^{\#}$ is kept constant, either set to zero or a value deduced in some other way.
3. For a quadrupolar nucleus the transverse relaxation rate, $R_2$, is obtained from **Eq. 3** if the correlation time and quadrupole coupling constant are known. Otherwise, it has to be measured directly from a solution where all the $^{43}Ca^{2+}$ ions are protein bound. $R_2$ can be used to calculate the product $\chi^{2}{}^{*}\tau_c$. To resolve this product into its to components we need a separate determination of $R_1$.
4. After obtaining $t_c$ at one temperature it can be calculated at any other temperature assuming the following equation to hold:

$$1/\tau_c = (kT/h)\exp(-\Delta G_0/RT) \quad (5)$$

5. The temperature dependence in the binding constant(s) is normally neglected because the experiments routinely are run under conditions such that the binding sites will be saturated at all temperatures. It is, however, no problem to include temperature-dependent binding constants if they have been obtained by other means.
6. The comparison between calculated and experimental spectra can be made with various degrees of sophistication, from comparison by eye to a full least-squares calculation made by the computer.

A warning may be appropriate at this point. Westlund and Wennerström ***(26)*** have shown that, outside the region $\omega_0\tau_c < 1.5$, even for a case where all observed nuclei are bound to the protein and with no chemical exchange, the temperature dependence of $\tau_c$ can give rise to a temperature dependent linewidth surprisingly similar to the one caused by exchange, **Fig. 1**.

## 3. Applications

Most of the $^{43}Ca$ NMR studies have been devoted to the EF-hand family of proteins, mainly calmodulin ***(2,3,27–31)*** and calbindin $D_{9k}$ ***(21–24,32)*** and also troponin C ***(2,16,28)***, S100 ***(33,34)***, and parvalbumin ***(2)***. In these studies, the experiments have been performed in different ways in order to extract various kinds of information and will therefore be used here as examples to show what can be obtained with $^{43}Ca$ NMR studies.

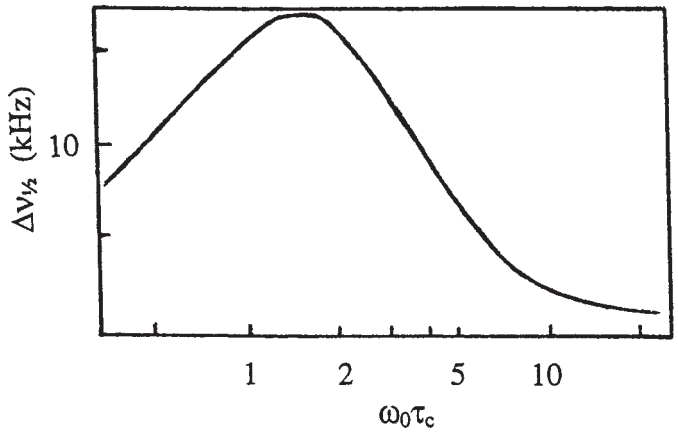


Fig. 1. The solid line shows the linewidth, $\Delta\nu_{1/2}$ for an I = 5/2 nucleus as a function of $\omega_0\tau_c$. Note the striking similarity to a temperature dependence considering that a factor of 10 in $\omega_0\tau_c$ corresponds to approx 70°C. Redrawn from **ref. *26***.

For a protein with high affinity for calcium, it is, in principle, possible to work under such conditions that all, or nearly all, $Ca^{2+}$-ions are bound to the protein and then a $^{43}Ca$ NMR study can give direct information regarding the binding site(s). Determination of the relaxation rates, $R_1$ and $R_2$, will make the calculation of the correlation time, $\tau_c$, as well as the quadrupole coupling constant, $\chi$, of the $Ca^{2+}$-ion possible by using **Eqs. 2** and **3**. The quadrupole coupling constant contains information regarding the symmetry of the calcium-binding site and the obtained correlation time will indicate whether the calcium ion has any mobility inside the binding site on the ns time scale or faster. $R_1$ can be determined with the standard inversion recovery pulse sequence. This is exemplified in **Fig. 2** for the $TR_1C$ fragment of calmodulin ***(20)*** and $R_2$ is most readily obtained directly from the linewidth, $R_2 = \Delta\nu_{1/2}/\pi$. As aforementioned, it is better to determine $R_2$ also using pulse techniques, however, as far as I know that has not been done. Applied to calmodulin, troponin C, and parvalbumin, it was found that the correlation times obtained agreed well with the correlation time for the macromolecule itself, showing that there is no fast motion of the calcium ion inside the binding sites. The quadrupole coupling constant was similar for all three proteins, 1.1–1.3 MHz ***(2)***. For two α-lactalbumins and two lysozymes, it was also found that the correlation time agrees reasonably well with that expected for the whole protein ***(6)***. The quadrupole coupling constants were found to be significantly smaller than for the EF-hand proteins, 0.7–0.8 MHz, showing that these sites are more symmetrical.

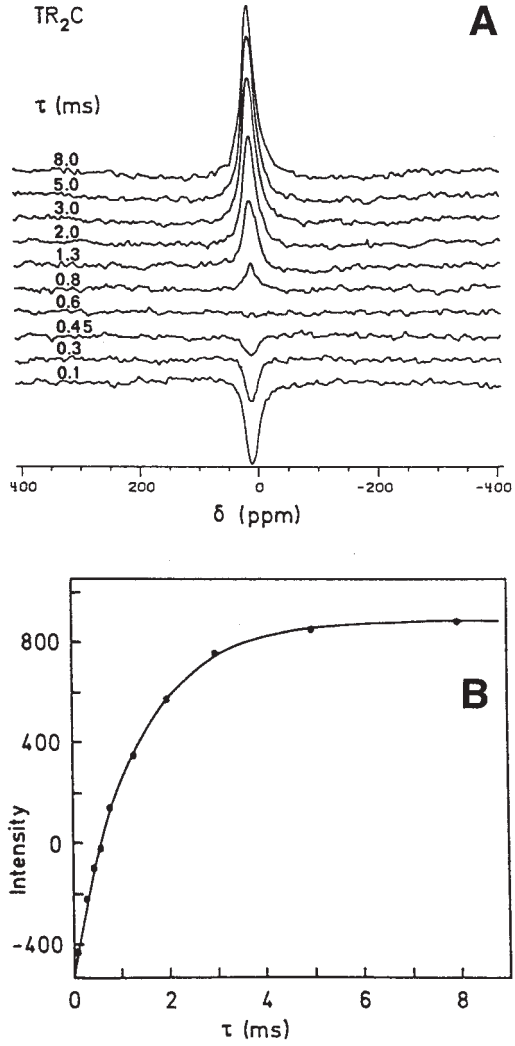


Fig. 2. Measurement of the relaxation rate $R_1$ for a $^{43}$Ca NMR signal from calcium bound to the *C*-terminal half of calmodulin. **(A)** $^{43}$Ca NMR spectra at various relaxation delay times in the inversion recovery experiment. **(B)** Intensity as a function of delay time. • Experimental points; solid line, best fit curve $R_1 = 770\ s^{-1}$. Reproduced with permission from **ref. *20***.

When the binding affinity for calcium is sufficiently high to have essentially no free calcium in the solution at nonsaturating concentrations, it will not be possible to determine the binding constant in a simple titration experiment in the same way as it will be for weaker binding (*see* next paragraph). However, if

it is possible to find a calcium chelator with a binding constant for calcium similar to that of the protein, it is possible to determine the binding constant in a competition experiment ***(6)***. By using an equimolar concentration of chelator and protein $Ca^{2+}$ will be evenly distributed between chelator and protein when they have the same affinity for $Ca^{2+}$. In **ref. *6***, the chelator affinity was fine-tuned to that of the protein (lysozymes and α-lactalbumin) by changing pH, **Fig. 3**. Using ethylenediaminetetracetic acid (EDTA) as chelator, it was in this way possible to determine $Ca^{2+}$-binding constants from $1 \times 10^6$ to $2 \times 10^7$ $M^{-1}$. These values are two to three orders of magnitude above what can be reached by the direct titration method, but are in the same range as those determined by using fluorescent chelators (*see* Chapter 2, this volume).

When the dissociation constant for calcium binding to a protein is close to the concentration used in the experiments, it can be determined from a titration experiment. However, if the calcium off-rate is not fast, compared to the relaxation rate of the bound ion, care has to be exercised to take this into account ***(25)***. For phospholipase $A_2$ ($PLA_2$) and prophospholipase $A_2$ ($PPLA_2$) the effect on exchange was taken into account by simultaneously fitting the data from a temperature dependence experiment and a concentration dependence experiment ***(17)*** (**Fig. 4**). This will result in the determination of both binding constant and kinetics of the calcium ion. Preferably the experiment should be executed on a single sample in the following way. (1) Start with as low a calcium concentration as possible for obtaining a reasonable spectrum in a reasonable time and titrate until the linewidth has been reduced to approximately half of the initial value. At this point, the protein should preferably be saturated with calcium to such a degree that a temperature-dependence in the binding constant has only a minor effect. (2) Perform the temperature dependence experiment on the same sample. (3) Repeat the last experiment in the titration experiment; and (4) complete the titration. If the protein is not thermostable, as will be seen from the nonidentity of the repeated experiment, separate samples for the titration and temperature dependence have to be used. When all exchange rates are fast on the NMR time scale, it is more straightforward to determine the binding constant directly from a concentration dependence in the linewidth. The temperature dependence will not result in any useful information in this case.

As shown in **refs. *6*** and ***17***, $^{43}Ca$ NMR can also be used to determine $pK_a$ values of residues involved in the binding site. For the case of strong binding and slow exchange, the $pK_a$-value can be calculated from the change in the intensity of the NMR signal from bound ions as a function of pH ***(6)***. For the case of weaker binding and fast exchange, the pH dependence in the line width can be used to calculate the $pK_a$-value ***(17)***. Without taking the effect of the calcium binding into account, this treatment will of course result in apparent $pK_a$-values.

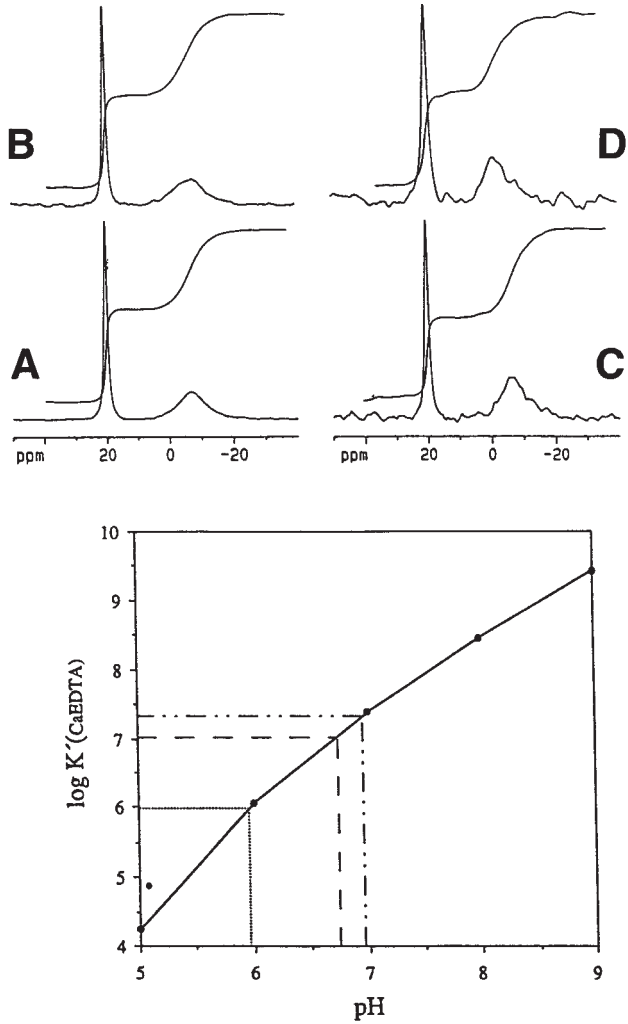


Fig. 3. **(A)** $^{43}$Ca NMR spectra of **(A)** bovine a-lactalbumin at pH 7.0; **(B)** human α-lactalbumin at pH 7.0; **(C)** equine lysozyme at pH 6.0; **(D)** pigeon lysozyme at pH 6.8. **(B)** Elucidation of dissociation constants ($K_D$) for the above for proteins using the pH dependence of the affinity of EDTA for $Ca^{2+}$. (····) Equine lysozyme, $K_D = 1 \times 10^6$ *M*; (– –) pigeon lysozyme, $K_D = 1 \times 10^{-7}$ M; (–· –) bovine and human α-lactalbumin, $K_D = 5 \times 10^{-8}$ *M*. Reproduced with permission from **ref. *6***.

However, when the calcium affinity is known the effect from the calcium binding can be taken into account to obtain a "true" $pK_a$-value by using **Eq. 6**.

$$\Delta\nu_{1/2}^{obsd} = \Delta\nu_{1/2}{*}K[E]/(1 + K[E]) \quad (6)$$

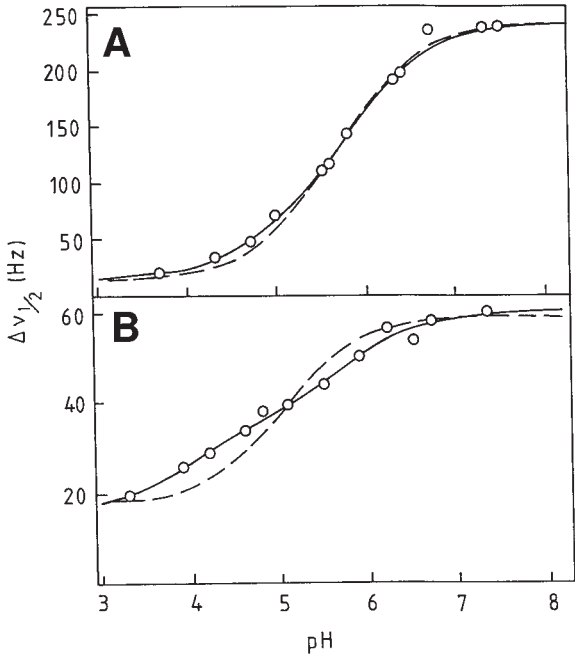


Fig. 4. Dependence of the $^{43}Ca$ NMR linewidth on pH at 23°C; **(A)** 0.98 m*M* phospholipae $A_2$ and 1.1 m*M* $^{43}Ca^{2+}$; **(B)** 1.1 m*M* phospholipase $A_2$ and 11.5 m*M* $^{43}Ca^{2+}$. The dashed curves are calculated using **Eq. 6** and the solid curves are calculated taking two binding sites into account. Reproduced with permission from **ref. *17***.

where

$$[E] = [E_{tot}]/(1 + K[Ca^{2+}] + 10^{pKa\text{-}pH})$$

and

$$[Ca^{2+}] = [Ca_{tot}]/(1 + K[E])$$

**Equation 6** can also be extended to take more than one binding site or $pK_a$ into account (**Fig. 4**) ***(17)***.

$^{43}Ca$ NMR has also been used to study the effect of drug binding on the calcium-binding properties of proteins ***(3,29,30)***. This is a clear example for the need of care in the interpretation, because the reduced line broadening on TFP addition was initially interpreted as a competition between $Ca^{2+}$ and TFP ***(30)***. It was later shown that the reduced line broadening is a result of a reduced exchange rate between free and bound $Ca^{2+}$-ions ***(29)***. This was shown quite elegantly with the use of shift reagents, whereby the $^{43}Ca$ NMR resonances from free and protein bound ions could be shifted apart, **Fig. 5**.

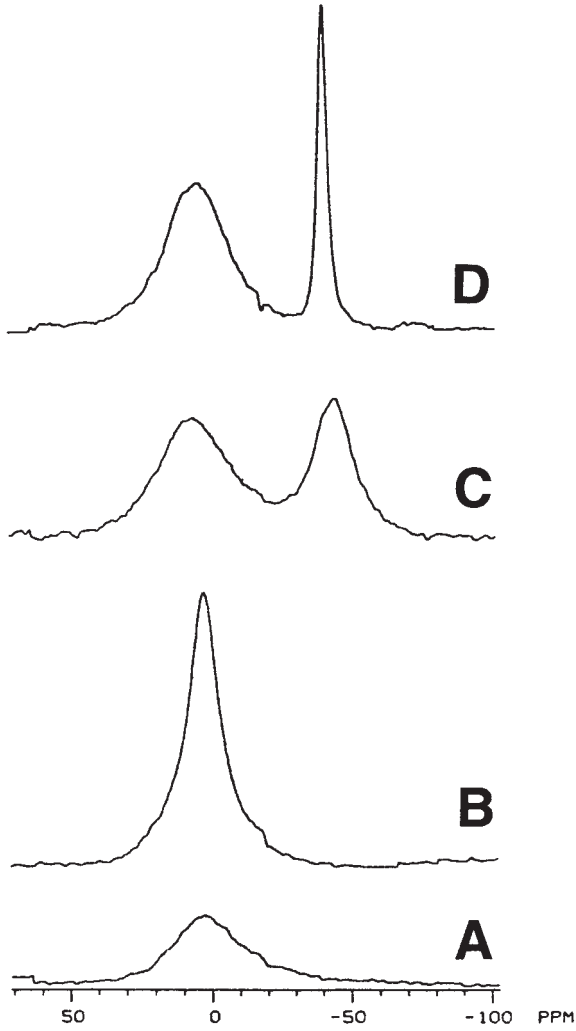


Fig. 5. $^{43}$Ca NMR spectra at 24.3 MHz obtained for a sample containing 1 m*M* bovine testis calmodulin, 100 m*M* Na $ClO_4$ at pH 7.0 with the following additions **(A)** 2.0 equ. $Ca^{2+}$; **(B)** 6 equ. $Ca^{2+}$; **(C)** 6 equ. $Ca^{2+}$ and 1.2 m*M* $Dy(PPP)_2^{7-}$; **(D)** 6 m*M* $Ca^{2+}$, 1.2 m*M* $Dy(PPP)_2^{7-}$, and 3 m*M* TFP. Reproduced with permission from **ref. *29***.

As aforementioned, the $^{43}$Ca NMR spectrum will depend strongly on the rate of calcium exchange between free calcium ions and ions bound to a protein. When the temperature dependence of the line shape is followed as a function of temperature, information about the exchange can be obtained if the exchange rate is faster than the relaxation rate of the free ion and not much

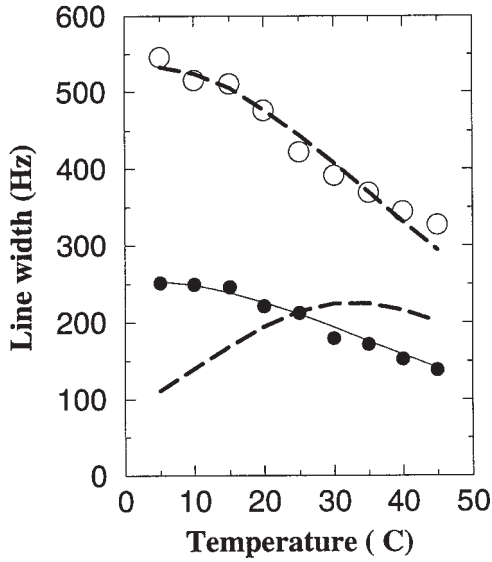


Fig. 6. $^{43}$Ca NMR linewidth as a function of temperature in the absence (open circles) and presence (filled circles) of 20 m*M* $Mg^{2+}$, pH 7.2 in 100 m*M* KCl. The upper-dashed curve shows the linewidth calculated using activation parameters from **ref. *20***. The lower dashed curve has been calculated with the same parameters with the population of free $Ca^{2+}$ adjusted to obtain the proper linewidth at 25°C. The best agreement with experimental linewidth, as shown by the solid curve was obtained using a quadrupolar coupling $\chi$ = 1.22 MHz and an off-rate at 25°C of 6600 $s^{-1}$ (compared to 500 $s^{-1}$ for the upper curve).

faster than the relaxation rate of the bound ions ***(19,35–37)***. It is thus possible to determine $Ca^{2+}$ off-rates from approx 10 $s^{-1}$ to approx $10^5$ $s^{-1}$. This means that $^{43}$Ca NMR can determine rates 100 times faster than those that can be obtained from stopped-flow studies. We have, in Lund, used the scheme outlined above on several occasions to study the $Ca^{2+}$ exchange. Most recently, we have applied it in a study of the Ca–Mg competition for the calcium-binding sites in the *N*-terminal domain of calmodulin ***(25)***. As shown in **Fig. 6**, it was possible to show that the exchange rate of $Ca^{2+}$ from the $Ca_1$ $Mg_1$-form of $TR_1C$ from calmodulin is much faster than from the $Ca_2$-form. This is completely in agreement with the strong cooperativity in calcium binding and the lack of cooperativity between calcium and magnesium.

I hope that we have been able to show that $^{43}$Ca NMR can be quite useful on many occasions when studying calcium-binding proteins. It is disappointing that so few groups have been considering the use of it to date. With the high sensitivity of modern NMR spectrometers, it should be possible to run $^{43}$Ca

NMR at sub-m*M* concentration without having probes specifically designed for this particular purpose. Even though the cost for enriched $^{43}$Ca may appear high, the total cost for a sample will almost always be dominated by the cost involved in preparation of the protein. Hopefully, in the future, more groups will consider the use of $^{43}$Ca NMR in their studies of calcium-binding proteins.

## References


1. Robertson, P., Hiskey, R. G., and Koehler, K. L. (1978) Calcium and magnesium binding to γ-carboxyglutamic acid-containing peptides via metal ion nuclear magnetic resonance. *J. Biol. Chem.* **253,** 5880–5883.
2. Andersson, T., Drakenberg, T., Forsén, S., Thulin, E., and Svärd, M. (1982) Direct observation of the $^{43}$Ca ions bound to proteins. *J. Am. Chem. Soc.* **104,** 576–580.
3. Shimizu, T. and Hatano, M. (1985) Magnetic resonance studies of trifluoperazine-calmodulin solutions: $^{43}$Ca, $^{25}$Mg, $^{67}$Zn, and $^{39}$K nuclear magnetic resonance. *Inorg. Chem.* **24,** 2003–2009.
4. Urry, D. W., Trapane, T. L., and Venkatachalam, C. M. (1982) Calcium binding to a calcifiable matrix: $^{43}$Ca NMR binding studies on the polypentapeptide of elastin. *Calcif. Tissue Int.* **34,** S41–S46.
5. Bouhoutsos-Brown, E., Pletcher, C. H., Nelsestuen, G. L., and Bryant, R. G. (1984) Prothrombin fragment 1-membrane interactions: a calcium-43 NMR study. *J. Inorg. Biochem.* **21,** 337–343.
6. Aramini, J. H., Drakenberg, T., Hiraoki, T., Nitta, K., and Vogel, H. (1992) Calcium-43 NMR studies of calcium-binding lysozymes and α-lactalbumins. *Biochemistry* **31,** 6761–6768.
7. Belton, P. S., Cox, I. Y., and Harris, R. K. (1985) Experimental sulphur-33 nuclear magnetic resonance spectroscopy. *J. Chem. Soc. Faraday, Trans 2* **81,** 63–75.
8. Ernst, R. R., Bodenhusen, G., and Wokaun, A. (1989) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford Science, Oxford, United Kingdom.
9. Bull, T. E., Forsén, S., and Turner, D. L. (1978) Nuclear magnetic relaxation of spin 5/2 and 7/2 nuclei including effect of chemical exchange. *J. Chem. Phys.* **70,** 3106–3111.
10. Halle, B. and Wennerström, H. (1981) Nearly exponential quadrupole relaxation. A perturbation treatment. *J. Magn. Reson.* **44,** 89–100.
11. Bull, T. E. (1972) Nuclear magnetic relaxation of spin-3/2 nuclei involved in chemical exchange. *J. Magn. Reson.* **8,** 344–353.
12. Drakenberg, T., Johansson, C., and Forsén, S. (1997) Metal NMR for the study of metalloproteins, in *Protein NMR Techniques* (Reid, D. G., ed.), Humana, Totowa, New Jersey.
13. Sandström, J. (1982) *Dynamic NMR Spectroscopy*, Academic, London.
14. McConnell, H. M. (1958) Reaction rates by nuclear magnetic resonance. *J. Magn. Reson.* **28,** 430–431.

15. Andersson, T., Drakenberg, T., Forsén, S., Wieloch, T., and Lindsröm, M. (1981) Calcium binding to porcine pancreatic phospholipase $A_2$ studied by $^{43}$Ca NMR. *FEBS Lett.* **123,** 115–117.
16. Andersson, T., Drakenberg, T., Forsén, S. and Thulin, E. (1981) A$^{43}$Ca NMR and $^{25}$Mg NMR Study of Rabbit Skeletal Muscle Troponin C. *FEBS Lett.* **125,** 39–43.
17. Drakenberg, T., Andersson, T., Forsén, S. and Wieloch, T. (1984) Calcium ion binding to pancreatic phospholipase $A_2$ and its zymogen: a $^{43}$Ca NMR study. *Biochemistry* **23,** 2387–2392.
18. Chiancone, E., Drakenberg, T., Teleman, O., and Forsén, S. (1985) Dynamic and structural properties of the calcium binding site of bovine serine proteases and their zymogens. *J. Mol. Biol.* **185,** 201–207.
19. Svärd, M., Drakenberg, T., Andersson, T., and Fernlund, P. (1986) Calcium binding to bone γ-carboxyglutamic acid protein from calf studied by $^{43}$Ca NMR. *Eur. J. Biochem.* **158,** 373–378.
20. Teleman, A., Drakenberg, T., and Forsén, S. (1986) Kinetics of $Ca^{2+}$ binding to calmodulin and its tryptic fragments studied by $^{43}$Ca–NMR. *Biochim. Biophys. Acta* **873,** 204–213.
21. Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmdén, K., et al. (1987) Structure-function relationships in EF-hand $Ca^{2+}$-binding proteins. Protein engineering and biophysical studies of calbindin $D_{9k}$. *Biochemistry* **26,** 6723–6735.
22. Brodin, P., Johansson, C., Forsén, S., Drakenberg, T., and Grundström, T. (1990) Functional properties of calbindin $D_{9k}$ mutants with exchanged $Ca^{2+}$ binding sites. *J. Biol. Chem.* **265,** 11,125–11,130.
23. Johansson, C., Brodin, P., Grundström, T., Thulin, E., Forsén, S., and Drakenberg, T. (1990) Biophysical studies of engineered mutant proteins based on calbindin $D_{9k}$ modified in the pseudo EF-hand. *Eur. J. Biochem.* **187,** 455–460.
24. Johansson, C., Brodin, P., Grundström, T., Forsén, S., and Drakenberg, T. (1991) Mutation of the pseudo-EF-hand of calbindin $D_{9k}$ into a normal EF-hand. Biophysical studies. *Eur. J. Biochem.* **202,** 1283–1290.
25. Malmendal, A., Linse, S., Evenäs, J., Forsén, S., and Drakenberg, T. (1999) The battle for the EF-hand: magnesium-calcium interference in calmodulin. *Biochemistry* **38,** 11,884–11,850.
26. Westlund, P.-O. and Wennerström, H. (1982) NMR line shapes of I = 5/2 and I = 7/2 nuclei. Chemical exchange effects and dynamic shift. *J. Magn. Reson.* **50,** 451–466.
27. Andersson, T., Drakenberg, T., Forsén, S., and Thulin, E. (1982) Characterization of the $Ca^{2+}$ binding sites of calmodulin from bovine testis using $^{43}$Ca and $^{113}$Cd NMR. *Eur. J. Biochem.* **126,** 501–505.
28. Drakenberg, T., Forsén, S., and Lilja, H. (1982) $^{43}$Ca NMR studies of calcium binding to proteins: interpretation of experimental data by bandshape analysis. *J. Magn. Reson.* **53,** 412–422.
29. Vogel, H. J., Andersson, T., Braunlin, W. H., Drakenberg, T., and Forsén, S. (1984) Trifluoperazine binding to calmodulin: a shift reagent $^{43}$Ca NMR study. *Biochem. Biophys. Res. Commun.* **122,** 1350–1356.

30. Shimizu, T., Hatano, M., Nagao, S., and Nozawa, Y. (1982) $^{43}$Ca NMR studies of $Ca^{2+}$-tetrahymena calmodulin complexes. *Biochem. Biophys. Res. Commun.* **106,** 1112–1118.
31. Shimizu, T. and Hatano, M. (1988) $^{43}$Ca nuclear magnetic resonance of $Ca^{2+}$-calmodulin solutions: effects of trifluoperazine and peptides. *Inorg. Chim. Acta* **152,** 257–260.
32. Vogel, H. J., Drakenberg, T., and Forsén, S. (1985) Structural differences in the two calcium binding sites of the porcine intestinal calcium binding protein: a multinuclear NMR study. *Biochemistry* **24,** 3870–3876.
33. Ogoma, Y., Shimizu, T., Hatano, M., Fujii, T., Hachimori, A., and Kondo, Y. (1988) $^{43}$Ca nuclear magnetic resonance spectra of $Ca^{2+}$-S100 protein solutions. *Inorg. Chem.* **27,** 1853–1855.
34. Ogoma, Y., Kobayashi, H., Fujii, T., Kondo, Y., Hachimori, A., Shimizu, T., and Hatano, M. (1992) Binding study of metal ions to S100 proteins: $^{43}$Ca, $^{25}$Mg, $^{67}$Zn and $^{39}$K n.m.r. *Int. J. Biol. Macromol.* **14,** 279–286.
35. Braunlin, W. H., Vogel, H. J., Drakenberg, T., and Bennick, A. (1986) A calcium-43 NMR study of calcium binding to an acidic proline-rich phosphoprotein from human saliva. *Biochemistry* **25,** 584–589.
36. Wahlgren, M., Dejmek, P., and Drakenberg, T. (1990) A $^{43}$Ca and $^{31}$P NMR study of the calcium and phosphate equilibria in heated milk solutions. *J. Dairy Res.* **57,** 355–364.
37. Wahlgren, M., Dejmek, P. and Drakenberg, T. (1993) Binding of $Mg^{2+}$ and $Ca^{2+}$ to β-casein $A^1$. A multinuclear magnetic resonance study. *J. Dairy Res.* **60,** 65–78.

# 19

# Exploring Familial Relationships Using Multiple Sequence Alignment

**Aalim M. Weljie and Jaap Heringa**

## 1. Introduction

Over the course of the past 30 yr, a multitude of calcium-binding proteins has been discovered that employ several unique structural motifs for calcium-ion binding. The first prominent family identified bound calcium via a helix-loop-helix structural motif, and was coined the EF-hand binding motif, as it occurs between the E and F helices of carp parvalbumin ***(1)***. Today, the EF-hand calcium-binding family is ubiquitous, with members implicated in varied roles such as calcium signaling cell response and calcium storage. More recently, other calcium-binding motifs such as those found in annexin repeats ***(2)***, C2 domain proteins ***(3)***, and EGF domain proteins ***(4)*** have been identified. **Table 1** summarizes the characteristic amino acid sequence properties of each of these domains as provided in the PROSITE protein motif recognition database ***(5)***.

Within each of these four families, there is a varying amount of primary sequence diversity that can be tolerated while maintaining calcium-binding properties. In addition, these motifs are often found within proteins as subunits, in conjunction with other functional protein domains. Such functionality is especially prevalent in eukaryotic EF-hand and C2-domain families, which often use calcium-binding in response to intracellular calcium-ion cascades as a means of cell signaling.

Given the vast number of known sequences and diversity of these protein families, the obvious and popular method to make sense of the data has been to examine these proteins via sequence alignments. Much effort has been given to classification and grouping of each of these families into subfamilies and/or ancestral relationships. In particular, the EF family has been subject to exten-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

**Table 1**
**Amino Acid Motifs for the Common $Ca^{2+}$-Binding Families**

| Family | Motif | PROSITE Accession |
|---|---|---|
| EF-hand | D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW] | PS00018 |
| C2 domain | [ACG]-x(2)-L-x(2,3)-d-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-d-[PA]-[FY] | PS00499 |
| Annexin | [TG]-[STV]-x(8)-[LIVMF]-x(2)-R-x(3)-[DEQNH]-x(7)-[IFY]-x(7)-[LIVMF]-x(3)-[LIVMF]-x(11)-[LIVMFA]-x(2)-[LIVMF] | PS00223 |
| EGF | [DEQN]-x-[DEQN](2)-C-x(3,14)-C-x(3,7)-C-x-[DN]-x(4)-[FY]-x-C | PS01187 |

sive analysis based on primary protein sequence ***(6–8)***, nucleic acid sequence ***(9)***, and other evolutionary ***(10)*** information. Also the annexins ***(11,12)***, C2 domain families ***(3)***, and EGF family ***(4)*** have previously been analyzed.

A researcher interested in a calcium-binding protein might have a number of different possible motivations to perform their own sequence alignments. For a novel protein, one may be interested to find out what sequence characteristics their protein shares with previously examined molecules and the degree of similarity to those other proteins or subfamilies. Apart from being a powerful tool to study primary sequence characteristics, sequence alignments can also be used to garner more complex information via secondary and tertiary structure predictions ***(13)***. In fact, the currently most sensitive methods in secondary structure prediction all rely on a multiple alignment as input. Also, evolutionary relationships and phylogenetic trees are normally established from reliable sequence alignments. Such avenues of analysis are particularly important, given the rate at which sequence information is becoming available, resulting in an ever-increasing gap between sequence and structure information.

Whereas much research has been concentrated on the analysis of sequence alignments, the most critical aspect of any such examination remains the quality of the alignment itself. An erroneous alignment will produce erroneous results and will also confuse the aforementioned reliant techniques. Manual alignments utilizing expert knowledge are the ideal way to go about sequence comparisons, and for small amounts of data this is certainly the method of choice. However, the obvious challenges of dealing with large amounts of sequence data have led to the development of many algorithms that use silicon, rather than human, computation power. Most multiple alignment methods

ensure that the quality of the final product remains high at reasonable levels of pairwise sequence similarities (>30% sequence identity). However, as the degree of sequence similarity decreases below 25%, it generally becomes difficult to separate a legitimate "signal," or sequence conservation, from "noise," or apparent similarities between motifs that are, in fact, unrelated and should not be matched. In-depth treatment of popular methods available for aligning sequences is available elsewhere (e.g., **ref. *14***) as this is beyond the scope of this chapter. The alignments described here are based on the dynamic programming (DP) algorithm as implemented in the program PRALINE ***(14)*** and the popular method CLUSTALW ***(15)*** as representative programs.

This chapter will concentrate on specific methods available for sequence alignment of calcium-binding proteins. One particularly important question to address in the examination of these proteins is which part of the polypeptide chain one wishes to align and the degree of similarity shared with other protein sequences under consideration. If the protein of interest belongs to a particular subfamily in which the topology is roughly the same, as is the case for the calmodulin and troponin C families, then a global alignment program would be ideal. Global-alignment methods align input sequences over their full length and can be confused if, for example, some sequences contain unrelated domains in addition to the calcium-binding domain. Therefore, if sequences are being compared of very different lengths and topologies (e.g., the entire family of EF-hand or all C2-domain proteins) and the region of interest is only the calcium-binding area, then a local alignment algorithm would be judicious. Both of these situations will be addressed in this protocol. Additionally, some techniques to create phylogenetic trees will be presented and specific results of sequence alignments will be given in the notes to provide simple, but concrete, examples (*see* **Subheading 4.2.**). Throughout, it will be assumed that the researcher has access to the World Wide Web, although it is certainly possible to perform alignments if sequence databases and alignment programs are available locally. A compilation of Web sites holding methods and/or databases important for this chapter is given in **Table 2**.

## 2. Materials

1. A sequence (or set of sequences) for alignment known as the target sequence (set).
2. Access to one or more sequence databases to search against, such as SWISS-PROT ***(16)***, EMBL ***(17)***, GENBANK ***(18)***, or PIR ***(19)***.
3. An alignment program suitable for aligning the sequences of interest such as PRALINE, or CLUSTALX ***(20)*** (*see* **Note 1**).
4. Any additional biochemical or structural evidence known about the sequence, such as metal-binding sites or secondary structure information.

**Table 2**
**Websites of Various Secondary Structure Prediction Methods and Related Services**

| Service | URL |
|---|---|
| GOR4 | http://absalpha.dcrt.nih.gov: 8008/gor.html |
| PHD[a] | http://www.embl-heidelberg.de/predictprotein/predictprotein.html[b] |
| Pred2ary | http://yuri.harvard.edu/~jmc/2ary.html |
| NNSSP[a] | http://dot.imgen.bcm.tmc.edu:9331/pssprediction/pssp.html |
| PREDATOR[a] | http://www.embl-heidelberg.de/cgi/predator_serv.pl |
| DSC[a] | http://bonsai.lif.icnet.uk/bmrn/dsc/dsc_read_align.html |
| Zpred[a] | http://kestrel.ludwig.ucl.ac.uk/zpred.html |
| Jpred | http://jura.ebi.ac.uk:88881 |
| COILS2 | http://www.isrec.isb-sib.ch/coils/COIL.S doc.html |
| SRS | http://srs.ebi.ac.uk |
| Entrez | http://www.ncbi.nlm.nih.gov). |
| PROSITE | http://www.expasy.ch/prosite |
| SWISS-MODEL | http://www.expasy.ch/swissmod/SWISS-MODEL.html |
| Modeller | http://guitar.rockefeller.edu/modeller/modeller.html |

[a]Method can also be run using the Jpred server.
[b]Mirror websites for PHD can be found here as well.

## 3. Methods

### 3.1. Retrieval of Sequences to be Used for Alignment

If the sequences to be used are already known and acquired, skip to **Subheading 3.1.3.**

1. Use the target sequence to search the general SWISSPROT database ***(16)*** for similar proteins. This comparison can be done either based on the sequence similarity using a local alignment tool such as BLAST ***(21,22)*** or using the protein name, if it is known. Alternately, sequences can be retrieved using a tool such as SRS ***(23)*** (*see* **Note 2**), which can flexibly search based on the name of the sequence or certain motifs. If this latter approach is used, it might be useful to extract sequences from multiple databases (such as PROSITE ***[5]*** and PFA*M* ***[24]***) to obtain a more complete data set for motif recognition.
2. Certain calcium-binding proteins will not show high similarity if the search is based on the complete protein sequence as the calcium-binding region may be relatively small, such as with EF-hand proteins containing other domains. In this case, the calcium-binding sequence stretch can be extracted and used as a "probe" to find other calcium-binding sequences. This approach can be fine-tuned by increasing or decreasing the length of the probe by adding or subtracting residues to the termini of the probe sequences as appropriate. This is particularly useful

when dealing with sequences containing repeating elements, such as most EF-hand proteins and annexins, in combination with a local alignment routine or special repeat recognition programs ***(25)*** to find multiple instances of a motif within the same sequence.

3. Typically, the set of sequences of interest acquired through sequence-similarity searches will include a number of sequences with high identity, such as the same protein from different species, after which the list progresses smoothly to those of lower identity. This can result in a prohibitively large sequence set, which might be reduced by extracting a representative and nonredundant subset of the sequences. One method that can be employed to accomplish such reduction is OBSTRUCT ***(26)***. This algorithm takes a user-specified set of sequences and range for the sequence identity (e.g., 0–50%) and then assembles a maximum possible subset of these sequences for which all pairwise sequence identities fall within the specified identity range. Specifying low ranges for the sequence identities (e.g., 0–30%) guarantees that a representative subset of nonrelated sequences will be constructed. Conversely, high-identity bounds (e.g., 85–100%) lead to the construction of a closely related (sub)family of sequences.
4. The sequences must be formatted such that the alignment program used will be able to understand the data. Common formats include PIR, FASTA, GCG/MSF, and CLUSTAL. This step is dependent on the program being used, and the documentation for each program will give the required format. Note that often during the sequences-search phase (such as retrieval with SRS) the output can be formatted appropriately without the need for manual intervention. *See* **Note 3** for some considerations when deciding which parts of the sequences to include. For PRALINE, the sequences should be provided in PIR or FASTA format, whereas CLUSTALX accepts any of the above formats.

### 3.2. Performing the Alignment

1. Generally, one must choose a 20 × 20 residue exchange-weights matrix, containing the likelihoods of each amino acid mutating into any other, and gap penalties prior to computing the actual alignment. Typically, a gap opening and a gap extension penalty are required, which control the ease for the algorithm to insert gaps during the alignment. A popular exchange matrix, also recommended for PRALINE and CLUSTALX, is the BLOSUM62 matrix ***(27)***, combined with a gap opening penalty of 12 and gap extension penalty of 1 (*see* **Note 4**). Again, each program will have different options and each algorithm will be different, hence, the documentation should be carefully read.
2. Run the alignment program using properly formatted sequences and appropriate options (such as writing alignment output into the MSF file format or printing the progressive alignment tree with PRALINE).

### 3.3. Analysis of the Resulting Alignment (see Note 5)

1. Secondary structure prediction of a target sequence is most reliable when the input is a good alignment of multiple sequences. The current state-of-the-art pre-

diction methods include PHD ***(28)***, PREDATOR ***(29,30)***, DSC ***(31)***, and NNSSP ***(32)***. Different underlying techniques have been explored by secondary structure predictors, such as neural networks (PHD), *k*-nearest neighbor algorithms (PREDATOR, NNSSP), machine learning (DSC), and statistical approaches (ZPRED ***[33]***, MULPRED ***[33a]***). All these methods use a multiple alignment as the basis for prediction and exploit the distribution of the amino acids at each alignment position. The SSPRED method ***(34)*** also relies on multiple alignments, but exploits the positional-residue exchange patterns of the amino acids, rather than their frequencies. The JPRED server at the EMBL-European Bioinformatics Institute (Hinxton, U.K.) (http://jura.ebi.ac.uk:88881/) conveniently runs methods such as PHD, PREDATOR, DSC, NNSSP, ZPRED, and MULPRED. The NNSSP method has to be activated explicitly when using the JPRED server, as it is the slowest of the ensemble. The server accepts a multiple alignment and will then predict a consensus secondary structure for the sequence on top of the alignment: Alignment positions showing a gap for the top sequence are deleted. A single sequence can also be given to the server. In the latter case, a BLAST-search is performed to find homologous sequences, which are subsequently multiply aligned using CLUSTALX and then processed with the user-provided single sequence on top in the alignment. If sufficient methods predict an identical secondary structure for a given alignment position, that structure is taken as the consensus prediction for the position. In case no sufficient agreement would be reached, the PHD prediction is taken. This consensus prediction is somewhat less accurate when the NNSSP method is not invoked or completed in the computer time slot allocated to the user.

2. Homology modeling is a relatively straightforward process once a good alignment is achieved. The target sequence must be aligned with a sequence or multiple sequences with three dimensional coordinates available. Programs useful for such analysis are MODELLER ***(35)***, Insight II (MSI Inc., San Diego, CA), and SWISS-MODEL ***(36)***.
3. The creation of phylogenetic trees is a popular way to analyze sequence alignments and to infer the associated likely modes of evolution. A widely used and acclaimed protocol for tree construction is neighbor joining (NJ) ***(37)***, which can conveniently be carried out by the program CLUSTALX. One of the advantages of the NJ technique is that it does not rely on a uniform rate of evolution for all the sequences, which, for example, is the case for the earlier UPGMA clustering technique ***(38)*** (*see* **Note 6**). When using CLUSTALX, the "exclude positions with gaps" option should be selected to obtain a reliable NJ tree. In order to assess the quality of the tree, a bootstrapped tree can also be created (*see* **Note 6**). In the CLUSTALX package, there are two accessory programs named NJPLOT and UNROOTED with which the NJ trees can be plotted. The first program requires knowledge of a "root," or common evolutionary ancestor that can be simulated by including a distantly related sequence (traditionally called an "outgroup") to the alignment. The latter program draws the trees without a root, and hence only provides relative evolutionary distances, without inference as to

the direction and origin of evolution. Some background information on phylogenetic analysis is given in **Note 6**.

## 4. Notes

### 4.1. Methodology Notes

1. In general, two basic classes of alignment programs have been developed: global and local methods. Global alignment programs attempt to align the sequences over their whole length, whereas local programs search only for the most conserved regions and leave the other parts of the sequences unaligned. The most effective alignment algorithm depends on the nature of the sequences to be aligned. Global algorithms produce the most accurate and reliable alignments when all the sequences in the data set are of similar length. However, when the sequences differ greatly in length, local alignment programs are often more successful at identifying the conserved regions.

   The two most-explored computational techniques for multiple-sequence alignment are Dynamic Programming (DP) ***(39,40)*** and, more recently, Hidden Markov Modeling (HMM) ***(41,42)***. The DP technique guarantees the finding of the highest scoring alignment determined from summing amino acid substitution scores minus any insertion/deletion penalties. HM*M* is a statistical approach, which is powerful if applied to sequence database searches. However, HM*M* approaches for multiple sequence alignment generally perform poorly when compared to other methods ***(43,44)***, mainly because of the inherently complex parameterization of the technique. As a consequence, the state-of-the-art multiple alignment methods are all based on the DP technique.

   Some recent evaluations of available multiple alignment techniques have been carried out ***(45)*** using a versatile database of benchmark alignments called BAliBASE ***(45)***. These showed the method PRRP ***(46)*** to be marginally the most accurate, closely followed by CLUSTALX, which is a much faster program. Virtually the same accuracy as CLUSTALX was attained by the PRALINE method when run on default parameters, not utilizing strategies such as profile-preprocessing or predicted secondary structure induced alignment. Other methods included in the assessment tests generally fell behind, such as the local alignment method DIALIGN ***(47)***, the HMM-based method HMMT ***(48)***, or the Gibbs-sampling method GIBBS ***(49)***. It must be stressed that DIALIGN was relatively successful in aligning sequence with very large insertions or deletions. A further discussion of alignment strategies and associated methods can be found in **Appendix I**.
2. SRS (Sequence Retrieval System) ***(23)*** is a powerful front-end program to access a large number of popular sequence databases. In addition to sequences, one can search based on motifs such as those within the PROSITE or PFAM databases. Care must be taken, however, as the algorithms used to create these databases may include false-positive results, and exclude false-negative ones. The documentation must be read carefully to establish which sequences are included/excluded (*see* **Subheading 4.2.** for addition practical considerations when using

these databases). Similar functionality to SRS is found in other tools such as the Entrez system.

3. In certain cases, it may not be useful to align the entire sequences such as with sequences of differing size or containing repeats. For example, if one uses a global alignment routine and has sequences that vary from less than 100 amino acids to greater than 1000, the program will generally distort the shorter sequences in attempting to create the global alignments. In this case, it might be advantageous to use a local multiple alignment method, such as DIALIGN ***(47)***. A local pairwise alignment tool such as BLAST can also be used to find local matches, and only these portions of the sequence of interest can then be aligned using a global alignment routine. This might be particularly appropriate in examining EF-hand or C2-domain proteins from different subfamilies, as these sequences also contain a variety of noncalcium-binding domains. The well-annotated SWISSPROT database will give the specific residues for each entry belonging to a particular domain, and this can be useful for extracting only the calcium-binding domains. It should be noted that for reliable alignments, smaller sequences (approx 30 amino acids) generally require slightly higher identities (>25 %) for reliable alignments, as compared to longer sequences. Repeats are also difficult as there may be certain sections of proteins that align better with incorrect topology (e.g., the *C*-terminal end of one sequence with the *N*-terminus of another).
4. The amino substitution weights are normally given as a 20 × 20 matrix, containing the weights for all possible amino acid exchanges. A multitude of such matrices are available, of which the most widely used are the classic PAM250 ***(50)***, the BLOSUM62 ***(27)***, and the Gonnet ***(51)*** matrix. The insertion/deletion penalties are used to decrease the alignment score when gaps need to be made to optimally match the two sequences. Normally, a pair of gap penalties is used, consisting of an opening penalty used once for each gap and an extension penalty applied to each incurring gap position. In practice, recommended values for the gap penalties are available for most amino acid exchange matrices (such as 10 and 1 for the PAM250 or 12 and 1 for the BLOSUM62 matrix, for gap opening and extension, respectively). For divergent sequences, lowered settings for the gap penalties could be attempted. Some alignment programs, such as CLUSTAL or MULTAL ***(52)*** employ optimized combinations of exchange matrices with associated gap penalties, which makes it less straightforward to vary gap penalties and/or try other residue exchange matrices.
5. Although multiple alignments can confer a wealth of information, numerous avenues of further analysis are open to the researcher in which an alignment plays a key role. It should be mentioned, however, that each mode of analysis often requires specific types of sequences to be included in the alignment together with associated information. For example, homology modeling requires sequences whose 3-D coordinates are known, and some types of phylogenetic trees require outlier sequences to be included. The successful generation of alignments useful in these applications may require the researcher to perform the alignment several times in an iterative fashion before conditions are fully satisfied, albeit care in

preparing the alignment is certainly worth the effort. During the course of such optimizations, it is also worthwhile considering the incorporation of external information garnered from biochemical or structural studies of the proteins in the alignment.

6. Three basic concepts exist to reconstruct the divergent evolution of a set of sequences: (a) parsimony ***(53,54)***; (b) distance ***(55)***; and (c) maximum likelihood methods ***(56)***. Parsimony methods try to reconstruct phylogenies by exploring the concept of minimum mutation. Distance methods are aimed at exploring a matrix containing all pairwise distances of a set of multiply aligned sequences. These methods also try to reconstruct the past using a minimalist approach; i.e., using as few evolutionary changes as possible. Maximum-likelihood methods attempt to construct the most probable tree based on the sequence data and a specific stochastic model of evolution. The additional information that can be expressed in the model, such as weighting of functional similarity and amino acid importance and the nature of insertions/deletions (e.g., **ref. *10***), can optimize the connectivity and branch lengths of the resulting tree. The package MOLPHY ***(57)*** is a speed-optimized maximum likelihood method that could be considered if evolutionary information is the most important analysis. In this chapter, we will restrict ourselves to distance-based methods as they are quick and can be applied easily onto the sequence data.

   The evolutionary relationships of a subject set of sequences are normally depicted in a tree. A tree is a special case of a connected graph where travel from each node to any other is possible through edges (branches) by only one path between any two such nodes. A tree contains interior and exterior (terminal) nodes. Normally, the input sequences are contemporary and referred to as the operational taxonomic units (OTUs). They correspond with the exterior nodes of the evolutionary tree, whereas the internal nodes represent ancestral sequences that must be guessed from the OTUs and the tree topology. The length of each branch connecting a pair of nodes may correspond to the estimated number of substitutions between two associated sequences. The minimal evolution hypothesis is that the "true" phylogenetic tree is the rooted tree (i.e., contains a node ancestral to all other nodes), which has the shortest overall length and thus comprises the lowest cumulative number of mutations.

   Distance methods derive a tree from a distance matrix, in which approximations are stored of all pairwise evolutionary distances between the tree constituents (i.e., the sequences). Distances can be obtained from sequence identities ***(55)*** or pairwise sequence alignment scores ***(58)***. Normally, an agglomerative cluster criterion is then used to construct the phylogenetic tree, reflecting the evolutionary information in the best possible way. Many clustering criteria have been introduced over the years and each has an underlying assumption of evolutionary dynamics. The first cluster method used in molecular sequence phylogeny was the UPGMA or group-averaging method ***(38)***. It takes the average value over all intergroup distances to measure the evolutionary distance between two groups of sequences and has the underlying assumption of identical mutation rates in all

lineages. The method of Saitou and Nei ***(37)*** (called NJ), is acclaimed by many workers in phylogenetic analysis. The method relies on a protocol of progressive pairwise joining of nearest sequences (each sequence being represented by a node) such that each time two nodes are joined, they are represented by an internal node. The two nodes selected at each step for joining are those that keep the overall tree length at a minimum. The NJ method has the advantage over the UPGMA technique in that it does not use the evolutionary (dis)similarity among groups, but merely is a strategy to join sequences and calculate branch lengths without assumptions regarding the rate of evolution. A general advantage of distance methods over parsimony and maximum likelihood methods is that they usually are much less CPU intensive as they employ a fixed strategy to arrive at a final tree without the need to sample the complete and vast tree space.

The simplest way to calculate the distance between sequences is by using the percent divergence, which for two aligned sequences involves counting the number of nonidentical matches (ignoring positions with gaps) divided by the number of positions considered. When a multiple alignment is used, all positions containing a gap in any of the sequences are commonly ignored. The real evolutionary time between the divergence of two sequences depends on the speed of the evolutionary clock, a matter of ongoing controversy. Even under a uniform clock, sequence identity as a measure of distance underestimates the real number of mutations. Certainly in diverged sequences there is an increasing chance that multiple substitutions have occurred at a site. The greater the divergence, the more the evolutionary times are underestimated. Kimura ***(59)*** corrected for this effect by curve fitting such that the corrected evolutionary time from the distance $K$ (percent divergence divided by 100) is given by $corrected\ K = -\ln(1.0 - K - K^2 / 5.0)$. The formula applies to cases with a reasonably uniform evolutionary clock and sequence identies from 15% and fits the data well from identities higher than 35%. It is good practice to start tree-building routines from such corrected sequence distances, which can be done by the PHYLIP package ***(60–62)***.

Because evolutionary trees may be a result of local traps in the search space, it is important to estimate the significance of a particular tree topology and associated branch lengths. Felsenstein ***(63)*** introduced the concept of bootstrapping, which involves resampling of the data such that the alignment positions are randomly selected and placed in some order and then a tree is generated by the original method. This process is repeated a statistically significant number of times. Comparison of frequencies at which the $N$-3 internal branches ($N$ is the number of sequences) occur in the original tree and those from bootstrapping allow probability estimates of significance. It is common practice to consider frequencies higher than or equal to 95% as supportive for the occurrence of an original tree branch in the bootstrapped trees. In this way, bootstrapping tests the stability of groupings given the data set and the method, thus lessening the chance of incorrect tree structures caused by conservative and/or back mutations. Bootstrapping can be performed for any method that generates a tree from a mul-

**Table 3**
**Summary of Sequence Retrieval Statistics for Calcium-Binding Domains (Number of Sequences)**

| | Pfam | PROSITE | After 50% cut-off |
|---|---|---|---|
| EF-hand | 790 | 479 | 188 |
| C2 domain | 109 | 50 | 60 |
| Annexin | 171 | 54 | 37 |

tiple alignment, although it must be kept in mind that the *biological* significance of a particular tree is not addressed.

A researcher wishing to derive a phylogeny from a given set of protein sequences, should first seek one or more sensitive multiple alignment routines to construct a good alignment. If secondary or tertiary structures are known, the information should be applied to check and enhance the alignment. A variety of phylogenetic methods should then be used (preferably including the NJ method) in conjunction with bootstrapping and the results compared carefully for consistency. Paradoxally, many wrong trees that can be derived from a particular sequence set will lead to far more interesting phylogenies than the one correct tree.

### 4.2. Specific Practical Examples Using Calcium-Binding Proteins

In order to provide some practical insight into the methodology outlined above, example alignments of the EF hand, C2 domain, and annexin families are provided below. The results presented are not intended to provide conclusive results of any sort, but to demonstrate a general methodology, simple interpretations of results, possible pitfalls and ways to avoid them. The emphasis is on alignment of sequences with large identity differences, and on contrasting the performance of different alignment methods (exemplified here by the PRALINE and CLUSTALX algorithms) in the alignment of the various families.

#### 4.2.1. Methodology

The basic procedure used was as described in **Subheading 3.**, and this section is included to provide some example of the scope and procedures involved in these alignments. Sequences were extracted from the PROSITE and PFAM databases using the terms “EF hand,” “C2,” and “annexin” via the SRS interface. **Table 3** provides a summary of the number of sequences available from each. Because of the variant nature in sequence length and the presence of repeats, it was decided that the alignments should be performed for the calcium-binding regions only, with each calcium-binding site having its own sequence entry. For each of the alignments, the sequences in these two databases were

compared, and a final file was created with all unique sequences. This included all false-negative entries from the PROSITE database and selected ancestral calcium-binding regions to examine the manner in which these regions have "evolved" out of calcium binding, as well as provide a reasonably difficult test set with sequences of low homology. Finally, these sequences were subjected to a 0–50% identity cut-off using OBSTRUCT (*see* **Table 3**). The final set of sequences were aligned with both PRALINE and CLUSTALX using the BLOSUM62 matrix with gap opening and extension penalties of 12 and 1, respectively. The resulting alignments were analyzed by creating NJ trees in CLUSTALX, and visualizing the resultant files with the UNROOTED program. Finally, the alignments were recreated with outlier sequences in order to establish distant branches, and then bootstrapped trees were created. These trees were visualized with the NJPLOT program.

### 4.2.2. Results

As a result of space limitations and the large size of the EF-hand and C2-domain family alignments, representative sequences are shown here; and the complete alignments as well as examples of the NJ trees are available via the world wide web at http://groningen.bio.ucalgary.ca/cabp-alignments.html. In the following text, the analysis refers to the complete alignments, and references to representative figures will be made explicitly. Briefly, the key results from these sets of alignments are presented below.

1. EF-hand family: Both programs aligned the key residues for this family in a reasonable manner (*see* **Fig. 1**). The first Asp that coordinates calcium in the EF-hands was aligned well in both programs for the sequences showing strong canonical EF-hand characteriztics, as were the other key ion-binding residues for most sequences. As expected, difficulties arose with sequences that were very distantly related showing little homology, as PRALINE left the three least homologous sequences with very little overlap to the previous sequences, and CLUSTALX introduced large gaps in the majority of sequences to deal with the same problem. The computations were repeated with the six least homologous sequences removed, which greatly improved both alignments. As shown in **Fig. 1**, the matrix and gap penalties used with the ClustalX algorithm resulted in more gaps throughout the alignment, whereas PRALINE only opened a maximum of two in any sequence. The phylogenetic trees, which were created using these alignments were reasonably stable (*see* website), although many key columns were discounted during bootstrapping, which was based only on alignment positions without gaps since those would reduce the stability of the tree.

   This could be remedied by further manual aligning of the sequences or removing the sequences containing serious gaps that correspond to key residues in the EF-hand pattern of the remaining sequences. Alternately, a differ-

Methods in Molecular Biology: Calcium-binding Protein Protocols
- Multiple Sequence Alignment -

Figure 1 (A)

```
AACB_CHICK1   Q99001     -NEFRASFNHF-DRD--HSGTLGPEEFKACLISLGY---
TCH2_ARATH3   P25070     -SDLKEAFELY-DLD--GNGRISAKELHSVMKNLGE---
CDP2_MAIZE4   P49101     -STIREIISEV-DTD--NDGRINYEEFCAMMRGGMQ---
S105_HUMAN2   P33763     -SSIDDLMKSL-DKN--SDQEIDFKEYSVFLTMLCM---
EFH1_TRYCR2   P41048     -DELRSAFFHY-DKY--KTGYVTRKQFTELFATLGE---
TCH2_ARATH1   P25070     -DDIKKVFQRF-DKN--GDGKISVDELKEVIRALSP---
S10D_BOVIN1   P02633     --ELKGIFEKYAAKE-GDPNQLSKEELKLLLQTEFPS--
EFH1_TRYCR1   P41048     -HDALRVIGQ---SEPQNADELSFSDFLLLMTREVD---
FCA1_TRYCR1   P07749     -QRRIELFKKFDK-----NETGK-LCYDEVHSGCLEVLK
ACTB_DICDI2   P24005     ARLALEEVNKRIRAYEEEKARLTEESKIPGVKGLG----
CC23_HOMAM    P80364     HCLSQGLSGL-LSLGKLFRDSSWTLSKEELSRGVSQ---
P2C1_ARATH    P49597     ESAAADIVVVDISAGDEINGSDVTSEKKMISR-------
SCP1_BRALA4   P04569     --DVPAVYNVITDGGKVTFDLNRYKELYYRLLTS-----
QR1_COTJA1    P23499     ---YPVHWQFYQLDQHPVDRSLTHSELAPLRASLVP---
```

Figure 1 (B)

```
AACB_CHICK1         ----NEFRASFN-HF-DRDHS-G-T-LGPEEFKACLISLGY---
TCH2_ARATH3         ----SDLKEAFE-LY-DLDGN-G-R-ISAKELHSVMKNLGE---
CDP2_MAIZE4         ----STIREIIS-EV-DTDND-G-R-INYEEFCAMMRGGMQ---
S105_HUMAN2         ----SSIDDLMK-SL-DKNSD-Q-E-IDFKEYSVFLTMLCM---
EFH1_TRYCR2         ----DELRSAFF-HY-DKYKT-G-Y-VTRKQFTELFATLGE---
TCH2_ARATH1         ----DDIKKVFQ-RF-DKNGD-G-K-ISVDELKEVIRALSP---
S10D_BOVIN1         ----ELKGIFEK-YA-AKEGD-PNQ-LSKEELKLLLQTEFPS--
EFH1_TRYCR1         -------HDALR-VIGQSEPQ-NADELSFSDFLLLMTREVD---
FCA1_TRYCR1         ----QRRIELFK-KF-DKNET-G-K-LCYDEVHSGCLEVLK---
ACTB_DICDI2         ---ARLALEEVNKRIRAYEEE-KAR-LTEESKIPGVKGLG----
CC23_HOMAM          HCLSQGLSGLLS--LGKLFRDSSWT-LSKEELSRGVSQ------
P2C1_ARATH          ------ESAAADIVVVDISA--G-DEINGSDVTSEKKMISR---
SCP1_BRALA4         -----DVPAVYN-VI--TDG--GKVTFDLNRYKELYYRLLTS--
CB45_MOUSE2         ----KENKLHFR-AV-DPDGD-G-H-VSWDEYKVKFLASKG---
```

Fig. 1. Selected sequences from an alignment of various EF-hand regions as described in the text. **(A)** The resulting alignment from PRALINE, and **(B)** from ClustalX. Note that the order of the sequences in **(B)** has been modified to facilitate comparison with **(A)**. The first column is the sequence ID with the occurrence number appended if there were multiple EF-hands from the same sequence used in the complete alignment (e.g., AAB°CHICK1 is the represents the first EF-hand of chicken α-actin, and the second column in **(A)** is the accession code, followed by the alignment).

ent software package could be used that uses more statistically stringent algorithms such as the PHYLIP package ***(60–62)*** for creation of trees. In terms of the overall approach to sequence alignments of this family, an alternative might be to consider EF-hand pairs, which generally form subdomains, as opposed to individual EF-hands, which constitute reasonably short segments of these proteins. Such a tact might aid the analysis of low-homology

sequences as increasing the domain size might provide more "signal" and lower the amount of "noise" in these alignments.

2. C2-domain proteins: The C2-domain proteins present a difficult case for automatic global alignment algorithms as two topologies exist for this motif. Structural superpositioning of the two topologies shows that strand 8 from topology II is equivalent to strand 1 in topology I, while both strands are sequentially similar. Manual alignments demonstrate that sequences with the topology I fold overhang at the *N*-terminal end by one strand, and the other group overhangs at the *C*-terminus ***(3)***. Hence, in the current example, these overhangs were not considered, and the alignments encompassed the middle of strand 2 to the middle of strand 8 in topology I, and the middle of strand 1 to the middle of strand 7 from topology II. The C2 domains are found as single, double, and triple domains, hence the same strategy was employed as with the EF-hand alignments; individual C2 domains were extracted and then aligned. The results from **Fig. 2** demonstrate that both programs separated the secondary structure elements from sequences that were reasonably highly conserved. However, the PRALINE algorithm continued to keep secondary structure elements together as the homology decreased, with a few exception in the *C*-terminal region, whereas the CLUSTALX program had serious problems with the low homology sequences. On the other hand, the PRALINE alignment shows a greater number of gaps within the aligned β-strand regions, which might hamper their recognition in the absence of structural knowledge. Both alignments were submitted to the JPRED server to ascertain which alignment provides the best prediction (*see* **Fig. 2**).The JPRED server was able to predict strands 1, 2, 3, and 5 from both alignments reasonably well. Strand 6 was not predicted from either alignment, and only the PRALINE alignment saw strand 4, although it was shifted towards the *C*-terminus. The PRALINE alignment was also more consistent in matching the residues thought to be important in calcium-ion coordination. These alignment differences are significant if one is interested in including homologs of lower identity for subsequent analysis, especially when attempting to predict secondary or tertiary structure. The phylogenetic trees generated from both alignments were much more stable than those from the EF-hand alignments, presumably because of the

---

Fig. 2. (*see opposite page*) Selected sequences from an alignment of C2-domain regions as described in the text. **(A)** The resulting alignment from PRALINE, and **(B)** from ClustalX. Note that the order of the sequences in **(B)** has been modified to facilitate comparison with **(A)**. The first column is the sequence ID with the occurrence number appended if there are multiple C2-domains from the same sequence used in the complete alignment, and the second column in **(A)** is the accession code, followed by the alignment. In both alignments, residues of synaptotagmin known to have β-strand secondary structure are denoted by "S" above the alignments. Secondary structure predictions as performed by the JPRED server for the top sequence in the alignments (but based on the complete alignments) are given under each of the two alignments.

Figure 2 (A)

```
Observed SS                          SSSSSS SSSSSS          SSSSSSSS          SSS
PUB1_SCHPO1   Q92462    -----------DQLVVRIL-QALDLP-AKD--SNG-FSDPYVKIYLL---PDR-KKKFQT
SYT5_HUMAN1   O00445    ---------QSGQLLVGIL-QAMGLA-ALD--LGG-SSDPYVRVYLL---XDK-RRRYET
KPCG_BOVIN    P05128    -------------IHVTVG-EARNLI-PMD--PNG-LSDPYVKLKLI-PDPRN-LTKQKT
KPCG_MOUSE    P05697    ----PTFCDHCGSLLYGLVHQGMKCS-CCE--MNV-HRRCVRSVPSL-CGVDH-TER-RG
KPCE_HUMAN    Q02156    --LKIKICE-----AVSLKPTAWSLRHAVGPRPQTFLLDPYIALNVDDSRIGQTATKQKT
PIP1_YEAST    P40020    --VKIRILSTQLLPRLNDTSPSRNNTNSFVKVEFHTDDEPTMPISID-KGTRISATEAST
UN13_CAEEL2   P27715    -------------ITLTVL-CAQGLI-AKD--KTG-KSDPYVTAQVG-------KTKRRT
PIP2_HUMAN    Q00722    -------------LSITVI-SGQFLS-ERS-------VRTYVEVELF-GLPGDPKRRYRT
PERF_MOUSE    P10820    -------------LVVSNF-RAEHLWGDYT-----TATDAYLKVFF-------GGQEFRT
Y366_MYCGE    P47606    --------------KVLTIIDESFLKASK--------SIGYQNLKI-----QKGDVAQEV
KPC1_YEAST    P24583    -------------LTIGIT-AARDVD-HIQSPMFARKPESYVTIKI-----DDTIKART
BCR_HUMAN     P11274    -------------LNV-IVHSATGFKQS---------SNLYCTLEV--DSFGYFVNKAKT
UN13_CAEEL1   P27715    -------------LCITI--KKARLQGAVD------EFNSYVTVKL--------QTVKST
Predicted SS                           ssss s                  sssssss         sss

Observed SS       SSS S S SSSSS SSS              SSSSSS SSS      SSSSSSSSS
PUB1_SCHPO1       KVH-R-KTLNPIF-NETFQFS-VPLA-ELAQRKLHFS-VYDFDRFSRHDLIGQVVL----
SYT5_HUMAN1       KVH-R-QTLNPHF-GETFAFK-VPYV-ELGGRVLVMA-VYDFDRFSRNDAIGEVRV-P--
KPCG_BOVIN        RTV-K-ATLNPVW-NETFVFNLKP-G-DVERR-LSVE-VWDWDRTSRNDFMGAMS-----
KPCG_MOUSE        RLQ-L-EIRAPTS-DE---IH-ITVG-E--ARNL-----IPMDPNGLSDPYVKLKL----
KPCE_HUMAN        N--------SPAWHDEFVT------D-VCNGRKIELA-VFHDAPIGYDDFVANCT-----
PIP1_YEAST        KSS-QGNGFNPIWDAEVS----ITLK-DTDLTFIKFM-VISEE-----TQIASVCL----
UN13_CAEEL2       RTI-H-QELNPVW-NEKFHFECHNST-DR----IKVR-VWDEDNDLKSKL--RQKLTRES
PIP2_HUMAN        KLSPSTNSINPVWKEEPFVFEKILMP-ELAS--LRVA-VMEEG----NKFLGHRI-----
PERF_MOUSE        GVV--WNNNNPRW-TDKMDFENVLLS-T--GGPLRVQ-VWDADYGWDDDLLGSCD-----
Y366_MYCGE        ALSRFLNCLNKNEWNQ-VAFELKKYCENDVRAMISIV-LFIQDLIKKNDLFTFYS-----
KPC1_YEAST        K-----PSRNDRW-SEDFQ---IPVE---KGNEIEIT-VYDK----VNDSLIPVAI----
BCR_HUMAN         RVY--RDTAEPNW-NEEFE---IELEGSQTLRILCYEKCYNKTKIPKED-----------
UN13_CAEEL1       TVAVRGNL--PCW-EQEFIFET-----NRPDDGMVLE-LWAKGVLW-DKLIGVHY-----
Predicted SS      sss             ssssss           sssss ss
```

Figure 2 (B)

```
Observed SS           SSSSSSSSSSS             SSSSSSSS            SSSSSSS  SS
PUB1_SCHPO1       --DQLVVRILQALDLPAKD----SNGFSDPYVKIYLLPDRKK--K----FQTKVHRKTLN
SYT5_HUMAN1       QSGQLLVGILQAMGLAALD----LGGSSDPYVRVYLLXDKRR--R----YETKVHRQTLN
KPCG_BOVIN        ----IHVTVGEARNLIPMD----PNGLSDPYVKLKLIPDPRNLTK----QKTRTVKATLN
KPCG_MOUSE        --------PTFCDHCGSLLYGLVHQGMKCSCCEMNVHRRCVRSVPS-LCGVDHTERRGRL
KPCE_HUMAN        ----LKIKICEAVSLKPTAW---SLRHAVGPRPQTFLLDPYIALN----VDDSRIGQTAT
PIP1_YEAST        ----VKIRILSTQLLPRLNDTSPSRNNTNSFVKVEFHTDDEPTMPISIDKGTRISATEAS
UN13_CAEEL2       ----ITLTVLCAQGLIAKD----KTGKSDPYVTAQVGKTKRRTRTI---HQELNPVWNEK
PIP2_HUMAN        ------LSITVISGQFLSER---SVRTYVEVELFGLPGDPKRRYR----TKLSPSTNSIN
PERF_MOUSE        --------LVVSNFRAEHLW---GDYTTATDAYLKVFFGGQEFRT----GVVWNNNNPRW
Y366_MYCGE        ------------KVLTIIDESFLKASKSIGYQNLKIQKGDVAQEVALSRFLNCLNKNEWN
KPC1_YEAST        ----------------------LTIGITAARDVDHIQSPMFARKPESYVTIKIDDTIKA
BCR_HUMAN         -----------------LNVIVHSATGFKQSSNLYCTLEVDSFGYFVNKAKTRVYRDTAE
UN13_CAEEL1       --------------------LCITIKKARLQGAVDEFNSYVTVKLQTVKSTTVAVRGNL
Predicted SS           sssssss                 ssssss            ssssss

Observed SS       SSSSSS              SSSSSSSSSS      SSSSSSSSS
PUB1_SCHPO1       PIFNETFQFS---VPLAELAQRKLHFSVYDFDRFSRHDLIGQVVL-
SYT5_HUMAN1       PHFGETFAFK---VPYVELGGRVLVMAVYDFDRFSRNDAIGEVRVP
KPCG_BOVIN        PVWNETFVFN---LKPGDVERR-LSVEVWDWDRTSRNDFMGAMS--
KPCG_MOUSE        QLEIRAPTSDEIHITVGEARNLIPMDPNGLSDPYVKLKL-------
KPCE_HUMAN        KQKTNSPAWHD-EFVTDVCNGRKIELAVFHDAPIGYDDFVANCT--
PIP1_YEAST        TKSSQGNGFNPIWDAEVSITLKDTDLTFIKFMVISEETQIASVCL-
UN13_CAEEL2       FHFECHNSTDR-IKVRVWDEDNDLKSKLRQKLTRES----------
PIP2_HUMAN        PVWKE-EPFVF-EKILMPELAS-LRVAVMEEGNKFLGHRI------
PERF_MOUSE        TDKMDFENVLL-STGGPLRVQVWDADYGWDDDLLGSCD--------
Y366_MYCGE        QVAFELKKYCENDVRAMISIVLFIQDLIKKNDLFTFYS--------
KPC1_YEAST        RTKPSRNDRWSEDFQIPVEKGNEIEITVYDKVNDSLIPVAI-----
BCR_HUMAN         PNWNEEFEIELEGSQTLRILCYEKCYNKTKIPKED-----------
UN13_CAEEL1       PCWEQEFIFETNRPDDGMVLELWAKGVLWDKLIGVHY---------
Predicted SS                          ssssss
```

much larger length of sequences, which provides the opportunity for a greater number of consensus residues in the C2 family. This leads to many more gapless columns than found with the EF-hand alignments, which enhances the sampling during the bootstrapping calculations. Finally, both alignments resulted in phylogenetic trees that appropriately classify the sequences into subfamilies (*see* website).

3. Annexin proteins: The annexins constitute a heavily studied protein family from a phylogenetic point of view (e.g., **refs.** ***11*** and ***12***) partially because of their important biological function. Another reason stems from the fact that the annexin repeat is a reasonably common eukaryotic motif; however, the number of orthologs and paralogs found provides a considerable challenge in classification because of the lack of defined identity between proteins from the same species. For example, 10 human annexins have a mean amino acid identity of 49.8 ±/– 4.1% ***(12)***, and estimates of mutational rates suggest that between very different species (e.g., plants and animals) the identity should be much lower. Paradoxically, the annexin repeat itself is particularly interesting because the sequences are all highly homologous, similar in length with essentially no gaps or deletions, lending themselves well to alignment. The alignment of the annexin proteins by PRALINE and CLUSTALX are given in **Fig. 3**. The degree of similarity in terms of sequence length is immediately striking, and on further analysis the conservation of key residues is also dramatic. The NJ trees created with these alignments showed reasonable stability on bootstrapping; however, neither alignment lends itself well to classification of the annexins into appropriate subfamilies. Presumably such analysis is better suited for alignments based on the entire annexin protein, and not simply on the repeat itself. It should be noted that the most complete classification of subfamilies to date has been accomplished through a combination of DNA and protein analysis, in conjunction with maximum-likelihood methods for phylogenetic analysis ***(11)***.

### 4.3. Multiple Alignment Method

The problem of finding an optimal or highest scoring alignment of two sequences was solved three decades ago with the DP technique ***(39)***, which guarantees the finding of the highest scoring alignment determined from summing amino acid substitution scores minus any insertion/deletion penalties. The amino substitution weights are normally given as a 20 × 20 matrix, containing the weights for all possible amino acid exchanges. The insertion/deletion penalties are used to decrease the alignment score when gaps need to be made to optimally match the two sequences. Normally, a pair of gap penalties is used, consisting of an opening penalty used once for each gap and an extension penalty applied to each incurring gap position. However, when applied to more than two sequences, the calculation of the optimal alignment by multidimensional implementations of the basic dynamic programming algorithm for sequence pairs ***(64)***, becomes computationally unfeasible. Even with localized

Methods in Molecular Biology: Calcium-binding Protein Protocols
- Multiple Sequence Alignment -

Figure 3 (A)

```
AN11_COLLI   P14950    ---------GTQHRDLIRIMVSRHEVDMNEIKGYYKKMYGISLCQA-IMDELKGGYETILVAL
ANX2_BOVIN   P04272    ---------GVDEVTIVNILTNRSNEQRQDIAFAYQRRTKKELASA-LKSALSGHLETVILGL
ANXX_DROME1  P22465    ---------GTDEQEIIDVLVGRSNQQRQTIKAVYEAEFERDLVDD-LKDELGGKFEDVIVGL
ANXD_CANFA1  Q29471    LNKACKGM-GTDEAAIIEILSSRTSDERQQIKQKYKATYGKDLEEV-FKSDLSGNFEKTALA-
ANX6_BOVIN1  P79134    IKDAISGI-GTDEKCLIEILASRTNEQIHQLVAAYKDAYERELEAD-ITGDTSGHFRKMLVV-
ANX6_MOUSE1  P14824    ---------GSDKESILELITSRSNKQRQEICQSYKSLYGKDLIED-LKYELTGKFERLIVNL
ANX3_HUMAN   P12429    ---------GTNEDALIEILTTRTSRQMKDISQAYYTVYKKSLGDD-ISSETSGDFRKALLTL
ANXD_CANFA3  Q29471    --DAGDGRWGTDELAFNEVLAKRSHKQLRATFQAYQILIDKDIEEA-IEAETSGDLQKAYLTL
ANX7_XENLA3  Q92125    LYQAGEGKLGTDESSFNLVLASRSFPQLKAVAEAYARISKRDLLSV-IGREFSGYIEDGLK--
ANX1_RODSP   P24551    ---------GTDEETLIEILWTRSNQQIREITSVYREELKKDIAKY-QTSDTSGEFRDALLAL
ANXX_DROME2  P22465    ---------GTEEATLVEILCTKTNEEMAQIVAVYEERYQRPLAEQ-MCSETSGFFRRLLTLI
ANX4_BOVIN   P13214    ---------GTDEVKFLTVLCSRNRNHLLHVFDEYKRIAQKDIEQS-IKSETSGSFEDALLAI
ANX6_BOVIN6  P79134    --DTTSGDKSSLETRFMMILCTRSYPDLRRVFQEFVKMTNYDVEHT-IKKEMSGDVRDVFVAI
ANXX_DROME3  P22465    ---------GTDDATLIRIIVSRSEIDLETIKQEFERIYNRTLHSAVVDAETSGDYKRALTAL
ANX4_FRAAN1  P51074    ---------GTDEDALTRVIVTRAERDLRDIKEVYYKKNSVPLEQA-VAKDTSGDYKAFLLTL
ANX7_DICDI   P24639    ---------GTDEKEFIKILTSRSLPHIAAVASEYIKHHKKHSLIKAIDSEFSGSIKTGLIAI
ANXB_XENLA   P27006    ---------GTDVGKWITIMTERSTPHLQKVFERYKSYSPYDMEES-IKKEVKGDLENAFLNL
ANXX_DROME   P22465    ---------GTDEEVFNRIMSHASFPQLRLVFEEYKVLSGQTIEQA-IKHEMSDELHEAMMAI
ANX6_CHICK   P51901    ---------GTDERTLTRIMISRSEIDLLNIRGEFIDLFDKSLYQM-IEKD-SGDYCKALLAL
GIA1_GIALA   P17063    ---------AKDEVQIAFIASEYSAESREKIAKAYVASYGKELPDD-IKKALKGGSEESLLMD
```

Figure 3 (B)

```
ANX1_COLLI        ---------GTQHRDLIRIMVSRHEVDMNEIKGYYKKMYGIS-LCQAIMDELKGGYETILVAL
ANX2_BOVIN        ---------GVDEVTIVNILTNRSNEQRQDIAFAYQRRTKKE-LASALKSALSGHLETVILGL
ANXX_DROM1        ---------GTDEQEIIDVLVGRSNQQRQTIKAVYEAEFERD-LVDDLKDELGGKFEDVIVGL
ANXD_CANFA1       -LNKACKGMGTDEAAIIEILSSRTSDERQQIKQKYKATYGKD-LEEVFKSDLSGNFEKTALA-
ANX6_BOVIN1       -IKDAISGIGTDEKCLIEILASRTNEQIHQLVAAYKDAYERE-LEADITGDTSGHFRKMLVV-
ANX6_MOUS1        ---------GSDKESILELITSRSNKQRQEICQSYKSLYGKD-LIEDLKYELTGKFERLIVNL
ANX3_HUMAN        ---------GTNEDALIEILTTRTSRQMKDISQAYYTVYKKS-LGDDISSETSGDFRKALLTL
ANXD_CANFA3       --DAGDGRWGTDELAFNEVLAKRSHKQLRATFQAYQILIDKD-IEEAIEAETSGDLQKAYLTL
ANX7_XENLA3       LYQAGEGKLGTDESSFNLVLASRSFPQLKAVAEAYARISKRD-LLSVIGREFSGYIEDGLK--
ANX1_RODSP        ---------GTDEETLIEILWTRSNQQIREITSVYREELKKD-IAKYQTSDTSGEFRDALLAL
ANXX_DROM2        ---------GTEEATLVEILCTKTNEEMAQIVAVYEERYQRP-LAEQMCSETSGFFRRLLTLI
ANX4_BOVIN        ---------GTDEVKFLTVLCSRNRNHLLHVFDEYKRIAQKD-IEQSIKSETSGSFEDALLAI
ANX6_BOVIN6       --DTTSGDKSSLETRFMMILCTRSYPDLRRVFQEFVKMTNYD-VEHTIKKEMSGDVRDVFVAI
ANXX_DROM3        ---------GTDDATLIRIIVSRSEIDLETIKQEFERIYNRTLHSAVVDAETSGDYKRALTAL
ANX4_FRAA1        ---------GTDEDALTRVIVTRAERDLRDIKEVYYKKNSVPLEQAVAK-DTSGDYKAFLLTL
ANX7_DICDI        ---------GTDEKEFIKILTSRSLPHIAAVASEYIKHHKKHSLIKAIDSEFSGSIKTGLIAI
ANXB_XENLA        ---------GTDVGKWITIMTERSTPHLQKVFERYKSYSPYD-MEESIKKEVKGDLENAFLNL
ANXX_DROME        ---------GTDEEVFNRIMSHASFPQLRLVFEEYKVLSGQT-IEQAIKHEMSDELHEAMMAI
ANX6_CHICK        ---------GTDERTLTRIMISRSEIDLLNIRGEFIDLFDKSLYQMIE--KDSGDYCKALLAL
GIA1_GIALA        ---------AKDEVQIAFIASEYSAESREKIAKAYVASYGKE-LPDDIKKALKGGSEESLLMD
```

Fig. 3. Selected sequences from an alignment of annexin repeat regions as described in the text. **(A)** The resulting alignment from PRALINE, and **(B)** from ClustalX. Note that the order of the sequences in **(B)** has been modified to facilitate comparison with **(A)**. The first column is the sequence ID with the occurrence number appended if there are multiple annexin repeat sequences from the same protein used in the complete alignment, and the second column in **(A)** is the accession code.

searches around the main diagonal of the multidimensional search matrix ***(65)*** designed to lower computational efforts, no more than nine sequences of up to 300 residues can be aligned ***(66)***. In general, a method to find an optimal align-

ment requires a number of computational steps and memory allocations of at least the order of the product of the sequence lengths. Rigorous methods for the alignment of four or more sequences therefore cannot evaluate all possible matches, but attempt to approach the optimal alignment by considering only a small fraction of all possible comparisons through repeated use of pairwise sequence matching. To overcome the computational problems, various heuristic approaches have been developed leading to a large number of programs using different strategies.

Traditionally, the most popular approach has been the progressive alignment method ***(58,67)***, where a multiple alignment is built up gradually by aligning the closest sequences first and successively adding in the more distant ones. Most widely used methods thus work in an agglomerative way by aligning sequences, following an heuristically determined precalculated order, until all sequences are joined in a final multiple alignment. Typically, the initial step involves performing all pairwise comparisons between the sequences, and the resulting alignment scores are used to represent pairwise sequence similarities, from which a phylogenetic tree (or guide tree) is constructed. The alignment steps then begin with joining the most similar sequence pair, followed by a gradual joining of the sequences in the order dictated by the guide tree. During this alignment process, single sequence pairs become aligned, as well as single sequences with blocks of already aligned sequences or pairs of prealigned blocks.

A number of alignment programs based on this method exist, for example MULTALIGN ***(68)***, MULTAL ***(52)***, PILEUP ***(69)***, CLUSTAL ***(15,70)***, and PRALINE ***(14)***. All these programs use global DP to construct an alignment over the entire lengths of the sequences. They differ mainly in the method used to determine the order of alignment of the sequences and the way in which they represent sequence blocks. For example, MULTALIGN aligns sequences by adding them one by one using a simple linear order. MULTAL uses a fast sequential branching method to align the closest pairs of sequences first and then subsequently align the next closest sequences to those already aligned. PILEUP constructs a guide tree using the so-called unweighted pair-group method using arithmetic averages (UPGMA) ***(38)***. A consensus method is then used to align larger and larger groups of sequences according to the branching order of the tree. CLUSTAL uses the NJ algorithm ***(37)***, which is widely used in phylogenetic analysis, to construct a guide tree. The method PRALINE does not use a precalculated search tree, but reevaluates at each alignment step, which sequences or blocks of sequences should be joined and aligned.

Several new alignment algorithms have recently been developed, offering fresh approaches to the multiple-alignment problem. A common point of interest has been the application of iterative strategies to refine and improve the initial alignment. A local alignment approach is implemented in the DIALIGN

program ***(47)***, which constructs multiple alignments based on segment-to-segment comparisons rather than the residue-to-residue comparisons used previously. The segments are incorporated into a multiple alignment using an iterative procedure. The PRRP program ***(46)*** optimizes a progressive, global alignment by iteratively dividing the sequences into two groups which are subsequently realigned using a global group-to-group alignment algorithm. PRALINE ***(14)*** employs profile-preprocessing and secondary structure prediction to guide the alignments in an optionally iterative fashion.

The program SAGA ***(71)*** does not employ DP, but uses a genetic algorithm (GA) to select from an evolving alignment population the alignment, which optimizes, as an Objective Function (OF), the weighted sum of pairs as used in the MSA program ***(66)***. More recently, a measure of consistency between the considered multiple alignment and a corresponding library of CLUSTAL pairwise alignments was taken. This OF was developed for the COFFEE algorithm ***(43)***. As mentioned in **Subheading 4.1.**, Hidden Markov models (HMM) have also been attempted as statistical models of the primary structure consensus for a sequence family ***(41,42)***. The program HMMT ***(48)*** uses a simulated-annealing method to maximise the probability that an HMM represents the sequences to be aligned.

## References


1. Kretsinger, R. H., Nockolds, C. E., Coffee, C. J., and Bradshaw, R. A. (1972) The structure of a calcium-binding protein from carp muscle, *Cold Spring Harb. Symp. Quant. Biol.* **36,** 217–20.
2. Smith, P. D. and Moss, S. E. (1994) Structural evolution of the annexin supergene family. *Trends Genet.* **10,** 241–246.
3. Nalefski, E. A. and Falke, J. J. (1996) The C2 domain calcium-binding motif: Structural and functional diversity, *Protein Sci.* **5,** 2375–2390.
4. Handford, P. A., Mayhew, M., Baron, M., Winship, P. R., and Campbell, I. D., Brownlee, G. G. (1991) Key residues involved in calcium-binding motifs in EGF-like domains. *Nature* **351,** 164–167.
5. Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27,** 215–219.
6. Kawasaki, H. and Kretsinger, R. H. (1995) Calcium-binding *Proteins* 1:EF-hands. *Protein Profiles* **2,** 305–490.
7. Heizmann, C. W. and Hunziker, W. (1991) Intracellular calcium-binding proteins: more sites than insights. *Trends Biochem. Sci.* **16,** 98–103.
8. Nakayama, S., Moncrief, N. D., and Kretsinger, R. H. (1992) Evolution of EF-hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories. *J. Mol. Evol.* **34,** 416–448.
9. Kretsinger, R. H. and Nakayama, S. (1993) Evolution of EF-hand calcium-modulated proteins. IV. Exon shuffling did not determine the domain compositions of EF-hand proteins. *J. Mol. Evol.* **36,** 477–488.

10. Kawasaki, H., Nakayama, S., and Kretsinger, R. H. (1998) Classification and evolution of EF-hand proteins. *Biometals* **11,** 277–295.
11. Morgan, R. O. and Fernandez, M. P. (1997) Annexin gene structure and molecular evolutionary genetics. *Cell Mol. Life Sci.* **53,** 508–515.
12. Morgan, R. O. and Fernandez, M. P. (1997) Distinct annexin subfamilies in plants and protists diverged prior to animal annexins and from a common ancestor. *J. Mol. Evol.* **44,** 178–188.
13. Heringa, J., Frishman, D., and Argos, P. (1997) Computational methods relating proteins sequence and structure, in *Proteins: A Comprehensive Treatise*, vol. I (Allen, G. ed.), JAI Press, Greenwich, Connecticut, pp. 165–268.
14. Heringa, J. (1999) Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment. *Comput. Chem.* **23,** 341–364.
15. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.
16. Bairoch, A. and Apweiler, R. (1999) The SWISS-PROT protein sequence data bank and its supplement TREMBL. *Nucleic Acids Res.* **27,** 49–54.
17. Stoesser, G., Tuli, M. A., Lopez, R., and Sterk, P. (1999) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **27,** 18–24.
18. Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouellette, B. F., Rapp, B. A., and Wheeler, D. L. (1999) Genbank. *Nucleic Acids Res.* **27,** 12–17.
19. Barker, W. C., Garavelli, J. S., McGarvey, P. B., Marzec, C. R., Orcutt, B. C., Srinivasarao, G. Y., et al. (1999) The PIR-international protein sequence database. *Nucleic Acids Res.* **27,** 39–43.
20. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882
21. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.
22. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.
23. Etzold, T., Ulyanov, A., and Argos, P. (1996) SRS: information retrieval system for molecular biology data banks. *Methods Enzymol.* **266,** 114–128.
24. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D., and Sonnhammer, E. L. L. (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27,** 260–262.
25. Heringa, J. and Argos, P. (1993) A method to recognize distant repeats in protein sequences. *Proteins* **17,** 391–411.
26. Heringa, J., Sommerfeldt, H., Higgins, D., and Argos, P. (1992) OBSTRUCT: a program to obtain largest cliques from a protein sequence set according to structural resolution and sequence similarity. Comput. Appl. Biosci., **8,** 599–600.

27. Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89,** 10,915–10,919.
28. Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232,** 584–599.
29. Frishman, D. and Argos, P. (1996) Incorporation of long-distance interactions in a secondary structure prediction method. *Prot. Eng.* **9,** 133–142.
30. Frishman, D. and Argos, P. (1997) Seventy-five percent accuracy in protein secondary structure prediction. *Proteins* **27,** 329–335.
31. King, R. D. and Sternberg, M. J. E. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Prot. Sci.* **5,** 2298.
32. Salamov, A. A. and Solovyev, V. V. (1995) Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J. Mol. Biol.* **247,** 11–15.
33. Zvelebil, M. J., Barton, G. J., Taylor, W. R. and Sternberg, M. J. E. (1987) Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* **195,** 957.

33a. Cuff, J. A. and Barton, G. J. (1999) Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* **34,** 508–519.

34. Metha, P., Heringa, J., and Argos, P. (1995) A simple and fast approach to prediction of protein secondary structure from multiply aligned sequences with accuracy above 70%. *Prot. Sci.* **4,** 2517–2525.
35. Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815.
36. Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis* **18,** 2714–2723.
37. Saitou, N. and Nei, M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.
38. Sneath, P. H. and Sokal, R. R. (1973) *Numerical Taxonomy*. Freeman, San Francisco, California.
39. Needleman, S. B. and Wunsch, C. D. (1970) A general method applicable to the search of for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443–453.
40. Smith, T. F. and Waterman, M. S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* **147,** 195–197.
41. Baldi, P., Chauvin, Y., Hunkapiller, T., and McClure, M. A. (1994) Hidden Markov models of biological primary sequence information. *Proc. Natl. Acad. Sci. USA* **91,** 1059–1063.
42. Krogh, A., Mian, I. S., Sjölander, K., and Haussler, D. (1994) Hidden Markov models in computational biology. *J. Mol. Biol.* **235,** 1501–1531.
43. Notredame, C., Holm, L., Higgins, D. G. (1998) COFFEE: An objective function for multiple sequence alignments. *Bioinformatics* **14,** 407–422.

44. Thompson, J. D., Plewniak, F., and Poch, O. (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res.* **27,** 2682–2690.
45. Thompson, J. D., Plewniak, F., and Poch, O. (1999) BAliBASE: a benchmark alignment database for the evaluation of multiple sequence alignment programs. *Bioinformatics* **15,** 87–88.
46. Gotoh, O. (1996) Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. *J. Mol. Biol.* **264,** 823–838.
47. Morgenstern, B., Dress, A. and Werner, T. (1996) Multiple DNA and protein sequence alignment based on segment-to-segment comparison. *Proc. Natl. Acad. Sci. USA* **93,** 12,098–12,103.
48. Eddy, S. R. (1995) Multiple alignment using hidden Markov models, in *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology* (Rawlings, C., Clark, D., Altman, R., Hunter, H., Hengauer, T., and Wodak, S., eds.), AAAI Press, pp. 114–120.
49. Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F., and Wootton, J. C. (1993) Detecting subtle sequence signals: the Gibbs sampling strategy for multiple alignment. *Science* **262,** 208–214.
50. Dayhoff, M. O., Barker, W. C., and Hunt L. T. (1983) Establishing homologies in protein sequences. *Methods Enzymol.* **91,** 524–545.
51. Gonnet, G. H., Cohen, M. A., and Benner, S. A. (1992) Exhaustive matching of the entire protein sequence database. *Science* **256,** 1443–1445.
52. Taylor, W. R. (1988) A flexible method to align large numbers of biological sequences. *J. Mol. Evol.* **28,** 161–169.
53. Camin, J. H. and Sokal, R. R. (1965) Computer comparison of new and existing criteria for constructing evolutionary trees from sequence data. *J. Mol. Evol.* **19,** 9–19.
54. Eck, R. V. and Dayhoff, M. O. (1966) in *Atlas of Protein Sequence and Structure* Natl. Biomed. Res. Found., Silver Spring, Maryland.
55. Fitch W. M. and Margoliash E. (1967). Construction of phylogenetic trees. *Science* **155,** 279–284.
56. Felsenstein J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17,** 368–376.
57. Adachi, J. and Hasegawa, M. (1996) MOLPHY v. 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comp. Sci.* Monographs, 28, 1–150. Institute of Statistical Mathematics, Tokyo.
58. Hogeweg, P. and Hesper, B. (1984) The alignment of sets of sequences and the construction of phyletic trees: an integrated method. *J. Mol. Evol.* **20,** 175–186.
59. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, England.
60. Felsenstein J. (1989). PHYLIP — phylogeny inference package (version 3.2). *Cladistics* **5,** 164–166.
61. Felsenstein J. (1990) PHYLIP Manual version 3.3. University Herbarium, University of California, Berkeley, California.

62. Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266,** 418–427.
63. Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *J. Evol.* **39,** 783–791.
64. Murata, M., Richardson, J. S., and Sussman, J. L. (1985) Simultanious comparison of three protein sequences. *Proc. Natl. Acad. Sci. USA* **82,** 3073–3077.
65. Carillo, H. and Lipman, D. J. (1988) The multiple sequence alignment. *SIAM J. Appl. Math.* **48,** 1073–1082.
66. Lipman, D. J., Altschul, S. F., and Kececioglu, J. D. (1985) A tool for multiple sequence alignment. *Proc. Natl. Acad. Sci. USA* **86,** 4412–4415.
67. Feng, D. F. and Doolittle, R. F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25,** 351–360.
68. Barton, G. J. and Sternberg, J. E. (1987) A strategy for the rapid multiple alignment of protein sequences: confidence levels from tertiary tructure comparisons. *J. Mol. Biol.* **198,** 327–337.
69. Genetics Computer Group. Program manual for the GCG Package, Version 8. 575 Science Drive, Madison, Wisconsin.
70. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882.
71. Notredame, C. and Higgins, D. G. (1996) SAGA: sequence alignment by genetic algorithm. *Nucleic Acids Res.* **24,** 1515–1524.

# 20

## Structure Determination by NMR

*Isotope Labeling*

**Monica X. Li, David C. Corson, and Brian D. Sykes**

### 1. Introduction

Solution NMR spectroscopy is used widely to determine the structure of proteins. The size of the proteins that can be studied has increased dramatically in the past decade as advances in pulse sequences, probe design, and instrumentation has been made. One major contributing factor to these advances has been the ability to utilize $^{2}H$, $^{13}C$, and $^{15}N$ isotopically labeled proteins in residue assignment strategies. For modestly sized proteins, the assignments can be accomplished by standard homonuclear $^{1}H$ 2D methodology ***(1)***. As the size of the proteins exceeds 10 kDa, the NMR spectra become more crowded with overlapping signals. With $^{13}C$ and $^{15}N$ labeling, the heteronuclear experiments have allowed the spectra to spread into two, three, or four dimensions, thus increasing the resolution and decreasing the assignment ambiguities ***(2)***. Another problem accompanying increasing protein size is sensitivity loss as a result of line broadening because of the decrease in $^{13}C$ and $^{1}H$ $T_2$ relaxation times. The most significant contribution to $^{13}C$ $T_2$ relaxation is the strong dipolar coupling between the $^{13}C$–$^{1}H$ spin pairs. $^{1}H$ $T_2$ relaxation arises from proton–proton dipolar couplings. Replacement of $^{1}H$ by $^{2}H$ can increase the $T_2$ relaxation times significantly. Thus, incorporation of $^{2}H$ into large proteins has been widely used to improve the quality of spectra by a reduction in the number of peaks and concomitant narrowing of linewidths ***(3)***. In addition to structural determination, heteronuclear multidimensional NMR has also been widely used to study protein dynamics and interactions of these molecules ***(3)***.

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

In order for these experiments to achieve common use, it is important to be able to isotopically label a protein in an efficient and cost-effective manner and to purify it in good yield, both uniformly and specifically. Practically, if the protein under consideration for NMR studies can be cloned and expressed in *Escherichia coli*, uniform labeling with $^{15}N$ is relatively straightforward and inexpensive by using defined media containing [$^{15}N$, 99%] ammonium chloride or ammonium sulfate as the nitrogen source. Uniform (>95%) labeling with $^{13}C$ is also relatively straightforward by using defined media containing [$^{13}C_6$, 99%] glucose. [$^{13}C_6$, 99%] glucose is the most reliable method in terms of giving high yield and high $^{13}C$-incorporation. It is also much less expensive than commercially available $^{13}C$-enriched media. Replacement of [$^{13}C_6$, 99%] glucose by other reagents like [1,2–$^{13}C_2$, 99%] acetate is possible ***(4)***. Partial or complete aliphatic $^{2}H$ incorporation has been obtained by growth of *E. coli* in defined media containing certain percentage of $D_2O$ and backbone $^{1}H_N$ can be exchanged out by dissolving the sample in $D_2O$. In terms of efficient type-specific labeling, the host bacteria can be grown on a defined medium supplemented with one or more isotope-labeled amino acids or amino acid precursors.

In the past few years, we have enjoyed great success in isotope labeling the $Ca^{2+}$-binding protein, troponin C. Using the pET expression system and the isopropyl β-D-thiogalactopyranoside (IPTG) induction protocol of Studier et al. ***(5)***, we are able to efficiently produce tens to hundreds of milligrams of the labeled protein from liter-scale growths of host *E. coli*. These labeled proteins made it possible for us to determine the solution structures and dynamics of skeletal and cardiac troponin C in a variety of states ***(6)***. We will summarize the procedures in this chapter. The methods described here should be applicable to other $Ca^{2+}$-binding proteins with their proper expression system and bacterial host.

## 2. Materials

1. Expression Medium (Modified M9 Medium), (1 L):
   a. $NaH_2PO_4$ (6 g)
   b. $K_2HPO_4$ (3 g)
   c. NaCl (0.5 g)
   d. $(NH_4)_2SO_4$ (1 g) (*see* **Notes 1** and **2**)
   e. D-glucose (2–10 g) (*see* **Notes 3–5**)
   f. 1 *M* $MgSO_4$ (4 mL) (*see* **Note 6**)
   g. 1 m*M* $FeSO_4$ (2 mL) (*see* **Note 7**)
   h. Mineral Mixture (0.5 mL) (optional, *see* **Note 8**)
   i. Vitamins and trace elements mixture (10 mL) (*see* **Note 8**)
   j. Appropriate antibiotics
   k. pH = 7.5

   Dissolve $NaH_2PO_4$, $K_2HPO_4$, and NaCl in 1 L of double distilled (dd) $H_2O$, adjust pH to ~7.5, and divide into two 2-L flasks with 500 mL each. Stop flask with a

pad of folded cheesecloth and cover with a double layer of aluminum foil. Autoclave. After autoclaving, this can be stored a day or two at room temperature or for several days at cold room temperature. The other ingredients should be made up, sterilized separately, and added just prior to inoculation.

2. TY Medium, (1 L):
   a. Bacto Tryptone (16 g)
   b. Bacto Yeast Extract (8 g)
   c. NaCl (5 g)
   d. Appropriate antibiotics
   e. pH = 7.5

   Dissolve all ingredients except antibiotics in 1 L dd$H_2O$ and sterilize by autoclaving. Add antibiotics before inoculation.
3. NZ Medium, (1 L):
   a. NZ amine (10 g)
   b. Bacto Tryptone (5 g)
   c. NaCl (5 g)
   d. $MgSO_4$ (2.5 g) (*see* **Note 6**)
   e. Appropriate antibiotics
   f. pH = 7.5

   Dissolve all ingredients except antibiotics and $MgSO_4$ in 1 L dd$H_2O$ and sterilize by autoclaving. Add antibiotics and $MgSO_4$ before inoculation.
4. IPTG (Isopropyl β-D-Thiogalactopyranoside): Dissolve 120 mg into 8 mL of water and filter-sterilizes. This is enough for a 1-L prep.
5. Antibiotics:
   a. 5% ampicillin: Weigh out and dissolve in water. Filter sterilizes into a sterile container. Store in freezer. This stock can be used for up to ten days but breaks down over time or heating up above 55°C. Use 1 μL per 1 mL of medium.
   b. 2.5% chloramphenicol: Weigh out with a sterile spatula and dissolve into 99% ethanol in a sterile test tube or other sterile container. This solution cannot be sterilized by either autoclave or filter since it breaks down with heat and ethanol breaks down the filter membrane in filter-sterilizers. This stock can be kept frozen up to 6 mo. Use 1 μL per 1 mL of medium.
   c. Other antibiotics: Follow specific protocols for specific antibiotics.
6. Vitamins and Trace Elements Mixture (*see* **Note 8**): Weigh out 10 mg per 1 mL each of biotin, choline chloride, folic acid, niacinamide, D-pantothenic acid, and pyridoxal chloride and 1 mg per 100 mL of riboflavin in $H_2O$. All above can be mixed and made as a single batch, then filter-sterilized, or made individually and filters sterilized as desired. Store frozen and in the dark. Use 10 mL of above mixture per 1 L of medium

   25% of nicotinc acid in $H_2O$, filter-sterilizes separately. Use 2.5 mL per 1 L of medium.

   10% of thiamine in $H_2O$, filer sterilizes separately. Use 0.5 mL per 1 L of medium.

7. Mineral Mixture (*see* **Note 8**): A solution made of 0.2 g/L of any of the following ingredients: $CuSO_4$, $MnSO_4$, $ZnSO_4$, and $CaCl_2$. Filter sterilizes and store at 4°C. Use 0.5 mL for each 1 L of medium.
8. Equipment:
   a. A selection of autoclavable glassware including test tubes, flasks, beakers, Pasteur pipets, pipet tips for pipetmen, autoclave trays, and access to an autoclave.
   b. Equipment for filter sterilization of nonautoclavable solutions: such as Millipore stericups and sterifil filter units with 0.22-micro filters (500 mL and 1 L sizes), Millex GS filter units (0.22 micro) plus 10-mL syringes.
   c. Access to a vacuum line, gas outlet, Bunsen burner, bacterial loops, and Petri dishes. Access to temperature controlled incubators or warm room with shakers, cold room, or refrigerators. Access to a UV-Vis spectrometer.
   d. Expression system: an expression system capable of efficiently expressing the target proteins at high levels. A bacterial cell strain compatible with the expression system. The cell strain should be capable of efficient growth on limited medium. In expressing TnC, we use the pET expression system with the cell strain BL21 (DE3) *pLys*S.
9. Expression medium for $^2H$-labeling, 1 L contains:
   a. $NaH_2PO_4$ (12 g)
   b. $K_2HPO_4$ (6 g)
   c. $(NH_4)_2SO_4$ (1 g) stock solution in $D_2O$ (*see* **Notes 1** and **2**)
   d. D-glucose (2–10 g) stock solution in $D_2O$ (*see* **Notes 3–5**)
   e. 1 *M* $MgSO_4$ (8 mL) stock solution in $D_2O$ (*see* **Note 6**)
   f. 1 m*M* $FeSO_4$ (6 mL) stock solution in $D_2O$ (*see* **Note 7**)
   g. Appropriate antibiotics: stock solution in $D_2O$ filter-sterilized
   h. pH = 7.5

   Dissolve up $NaH_2PO_4$ and $K_2HPO_4$ in 500 mL of $D_2O$, adjust pH with NaOD and DCl and filter-sterilizes into a sterile 2-L flask (using a 500 mL Millipore filtration unit or equivalent). This should be done the same day or the day before the cells are to be grown. If this solution is stored at room temperature or cold room temperature, the medium must be warmed up to incubation temperature before introducing the cell culture. Other ingredients should be treated the same as aforementioned except in $D_2O$ and added just before use.

   $D_2O$: 99.9% $D_2O$ available from commercial sources.

   NaOD/KOD: Dissolve some NaOH or KOH in $D_2O$.

   DCl: Make a 2-*M* solution of HCl in $D_2O$

   IPTG: 120 mg in 10 mL $D_2O$, filter-sterilized.

   Vitamins and trace elements mixture: Same ingredients as before except in $D_2O$, filter-sterilized.

   Activated charcoal: Available from commercial sources. This is used to recycle $D_2O$.

   TY or NZ amine media: Makeup the required volume as aforementioned except in $D_2O$. Filter sterilizes.

## 3. Methods

### 3.1. Growth of Cells on Limited Media

Transform cells with the appropriate plasmid according to standard protocols (*see* **Note 9**). Plate out overnight at the correct temperature for your cells (usually 37°C). Inoculate five or six colonies from the plate of transformed cells into a 10-mL NZ-amine culture (*see* **Note 10**) containing the proper antibiotics in a 25-mL Erlenmeyer flask and shake at the proper temperature until the cells reach late-log phase, e.g., $OD_{600nm}$ = 1.0. Inoculate 1 L of expression medium with 5 mL of the NZ culture. Incubate at the proper temperature to mid-log phase, $OD_{600nm}$ = approx 0.5–1.0. Add IPTG to induce the cells and incubate for the required induction time (*see* **Note 11**).

### 3.2. Expression of Uniformly 15N-Labeled or 15N-Depleted Protein

Grow culture and express cells according to above methodology except that 99.9% $^{15}N$-$(NH_4)_2SO_4$ or $^{15}N$-$NH_4Cl$ (*see* **Note 1**) is used in the expression media as the sole nitrogen source for uniformly $^{15}N$-labeled protein (*see* **Note 12**). For $^{15}N$-depleted protein, use $^{15}N$-depleted [i.e., 99.99% $^{14}N$-$(NH_4)_2SO_4$ or $NH_4Cl$].

### 3.3. Expression of Uniformly 13C-Labeled or 13C-Depleted Protein

Grow culture and express cells according to above methodology except that 99.9% $^{13}C_6$ glucose is used in the expression media as the sole carbon source for uniformly $^{13}C_6$-labeled protein. For $^{13}C$-depleted protein, use $^{13}C_6$-depleted (i.e., 99.99% $^{12}C_6$-glucose).

### 3.4. Incorporation of Type-Specific Amino Acids

Grow culture and express cells according to above methodology. To the standard expression medium, add 20 common amino acids to the concentration of 0.1 g/L (*see* **Note 13**). These amino acids should be unlabeled except for the amino acid targeted, which should include the desired label. Of this, it is probably most efficient to add 0.033 g when making media, another 0.033 g at induction, and the remaining 0.033 g halfway through the induced period (*see* **Notes 14–17**).

### 3.5. Expression of 2H-Labeled and 2H,13C,15N-Labeled Proteins

#### 3.5.1. Conditioning of Cells to 80% $D_2O$ (see **Note 18**)

1. Transfer four or five colonies from a plate of freshly transformed cells into a sterile container containing 20 mL of NZ amine medium made in 20% $D_2O$/80% $H_2O$ with the proper antibiotics. This is incubated at 37°C (*see* **Note 19**).
2. When this culture has reached late-log phase ($OD_{600nm}$ = 1.0), inoculate 1 mL to a similar culture in 45% $D_2O$/55% $H_2O$ (*see* **Note 20**).

3. When the 45% $D_2O$/55% $H_2O$ culture reached mid- to late-log phase ($OD_{600nm}$ = 0.5–1.0), use these cells to inoculate a similar culture in 80%$D_2O$/20%$H_2O$.
4. When cells from **step 3** have grown to mid- or late-log phase, use these cells to inoculate a flask of M9 or modified M9 expression medium prepared in 80% $D_2O$/20% $H_2O$.
5. Grow **step 4** cells until late-log phase ($OD_{600nm}$ = 0.5–1.0), add 30% by volume of glycerol, and flash freeze. Store these cells at –80°C.

### *3.5.2. Plasmid Retention Test* (see **Note 21**)

1. Use one vial (1-mL aliquot) of frozen (at –80°C) conditioned cells prepared as above, to inoculate a sterile flask of 50 mL of NZ amine medium (or similar enriched media, such as TY media) with the required antibiotics.
2. Incubate until mid-log phase ($OD_{600nm}$ = 0.5–1.0).
3. Induce protein expression by IPTG.
4. After the required induction time, harvest cells and run a gel to check protein expression levels.

### *3.5.3. Protein Expression*

1. Use 1 mL (*see* **Note 22**) of frozen (–80°C) conditioned cells to inoculate a sterile container of 30–50 mL of the expression medium (note, it is recommended to use large amounts of glucose, 10 g/L in this step, *see* **Note 23**) in 80%$D_2O$/$H_2O$ as described in **Subheading 2.**
2. Incubate overnight or until late-log phase (*see* **Note 24**).
3. Use 30–50-mL cells (*see* **Note 24**) from **step 2** to inoculate 500 mL of expression medium in 99.9% $D_2O$ to achieve the highest levels of deuteration.
4. Grow until mid-log phase (*see* **Note 25**).
5. Induce by using IPTG in $D_2O$ and incubate for the appropriate time (*see* **Note 26**).
6. Harvest cells according to usual methods.

### *3.5.4. A Suitable Method for Recycling D2O* (see **Note 27**)

1. Remove visible debris by centrifuging and/or settling and/or filtering.
2. Using a common laboratory distillation apparatus, distill two or three times.
3. Mix with activated charcoal (1 g/100 mL works well) and stir for 20 min.
4. Remove charcoal by filtration. This $D_2O$ is ready for growing cells (*see* **Notes 28** and **29**).

## 4. Notes

1. $NH_4Cl$ can be used in place of $(NH_4)_2SO_4$, but the latter is slightly better because
   a. The chloride ion is of no use at all to most nonhalophilic bacteria like *E. coli*, although sulfate is essential;
   b. Considering cost vs stoichiometry, $(NH_4)_2SO_4$ will sometimes give a little more $^{15}N$ per dollar (depending on relative prices).
2. The $^{15}N$-$NH_4Cl$ or $(NH_4)_2SO_4$ can be mixed and autoclaved together with the phosphate buffers, but it is better to make it up separately and add it just before

cell growth because in the event that the starter culture does not grow properly, or for some reason, you decide not to go ahead with cell growth, the $^{15}N$-stock solution can be conveniently stored frozen for an indefinite period of time and used at a later date.

3. Glucose should not be autoclaved in the presence of salts and buffer because it can break down and form toxic complexes inhibitory to the metabolism of the cells. It is best to make it up separately, 2 g in 10 mL or 10 g in 50 mL dd$H_2O$, and filter-sterilizes it into a sterile container. This can be stored frozen indefinitely or at cold room temperature for several days.
4. Because glucose is the sole carbon source, it is important to have plenty available for the most quick and efficient cell growth, thus, for unlabeled $^{12}C$-glucose, up to 10 g/L or even 12 g/L can be used. However, for expensive $^{13}C_6$-glucose, it is important to use the least amount that gives the most efficient target protein expression. For preps in $H_2O$, 2 g/L has been found in our lab and others to give the most cost-effective result, i.e., most protein yield per gram of glucose used per liter prep. However, the yield per liter can be increased somewhat by adding extra glucose up to the level of 3 g/L, after which point adding extra is probably wasted. Please note, for deuterated preps, 2.5 g/L seems to give the highest yield per gram of glucose, whereas increasing glucose to 6 g/L gives greater amount of protein per liter prep.
5. When ordering $^{13}C$-labeled glucose, 6-$^{13}C$-glucose refers to glucose with the $^{13}C$ label in the number 6 position, whereas $^{13}C_6$-glucose refers to uniformly labeled glucose. Unlabeled, primarily $^{12}C$-glucose is simply called D-glucose (or dextrose), whereas $^{12}C_6$-glucose refers to $^{13}C$-depleted glucose.
6. Make 10–20 mL of $MgSO_4$ in an autoclaveable stoppered container and autoclave. This can be stored indefinitely in cold room or at room temperature, but it is better to make fresh every time.
7. $FeSO_4$ solution should be filter-sterilized. Autoclaving will result in forming $Fe(OH)_3$ and precipitation.
8. For *E. coli* strain BL21, vitamins and minerals are not absolutely essential because the cell has plenty of trace elements from the starter culture used to initiate growth and can make its own vitamins as required. For some other stains, some of these could be essential. Nevertheless, different researchers may favor adding different things to their media. It is recommended to do trial expressions to test if these ingredients help.
9. Freshly transformed cells from frozen stocks of plasmid are the most reliable and commonly used. Frozen stocks of transformed cells may work equally well, based on our experience.
10. TY medium can be used instead of NZ amine medium as an enriched starter culture. However, NZ amine is a less enriched and more strenuous medium so the cells should suffer less shock when introduced into the limited expression media.
11. The best induction time may be the same or may be longer than on enriched medium. It is strongly advised to try out a trial growth to check that the expres-

sion system works efficiently under these conditions and to familiarize the researcher with all the techniques involved. During this trial, it is also recommended to run a time course of target protein production against induction time.

12. If the only source of nitrogen is $^{15}$N-labeled, then the cell has to incorporate $^{15}$N into any proteins produced because no other source of nitrogen is available. For the techniques described herein, $(NH_4)_2SO_4$ or $NH_4Cl$ will be the primary source of nitrogen, with the only other source being the small amount available from the starter culture. This is why the amount of starter culture used for inoculation should be kept to minimum (1% or less). Mass spectroscopy analysis has shown that our methods result in protein with 96% or greater $^{15}$N incorporation. The same consideration applies to the expression of $^{15}$N-depleted, $^{13}$C-labeled, and $^{13}$C-depleted protein.
13. This results in a medium, which is much more enriched than M9, or a similar limited medium. Thus, it may be possible to grow strains that will not normally grow on limited media. Of course, proteins with uniformly $^{15}$N or $^{13}$C labeling cannot be expressed from this medium.
14. The efficiency of incorporation for any specific amino acid will vary according to amino acids metabolic pathway, particularly its catabolic pathway in a particular cell strain. Therefore, glycine, serine, cysteine, glutamic acid, and aspartic acid, which are used as metabolic precursors for a variety of pathways, tend to incorporate with poor efficiency, whereas threonine, valine, alanine, leucine, and isoleucine are incorporated fairly well. The $^{13}$C-labeled methyl group of methionine incorporates very well. For more information on the amino acid metabolism, refer to Muchmore et al. *(7)*.
15. *See* **Table 1**.
16. There are quite a few ways to vary the timing of addition of labels during growth. The major considerations here include:
    a. Adding label at the beginning of cell growth wastes label on cell proteins other than the target protein;
    b. However, without having available any of the targeted amino acid, the cell must induce the necessary enzymes required to make that particular amino acid. Thus, when the labeled amino acid is added, it will be in direct competition with unlabeled endogenous synthesized amino acids, which ends up diluting out the label anyway.
    c. Likewise, whereas adding the unlabeled version of the target amino acid at the beginning of the cell growth may cut off the induction of its synthetic pathway, in the end, the cell will still be making proteins from a mixed pool of labeled and unlabeled amino acid any ways. Thus, for maximal efficiency of label incorporation, it is probably best to add some of the labeled amino acid at the beginning of the cell growth and the remainder during induction, either at the beginning of the induction or at times intervals throughout.
17. The most efficient method to incorporate any given amino acid is to get an auxotrophic strain for that particular amino acid. However, that strain must be compatible with the expression system being used, which may not be readily available.

**Table 1**
**Taking Care of Specific Type**
**of Amino Acids in Type-Specific Labeling Preps**

| Amino acid | Keep separate | Mix together | Filter sterilize only | Autoclave or filter-sterilize | Store in fridge | Store at room temperature |
|---|---|---|---|---|---|---|
| Trp | √ | | √ | | √ | |
| Tyr | √ | | √ | | √ | |
| Asp, Glu | | √ | √ | | √ | |
| Ala, Leu, Ile, Val | | √ | | √ | | √ |
| Met | √ | | | √ | √ | |
| Phe, Pro, Ser, Thr | | √ | | √ | √ | |
| Arg, Lys | | √ | | √ | √ | |
| Gly | √ | | | √ | √ | |
| His | √ | | √ | | √ | |
| *Asn, Gln, Cys | | | | | | |

*Bacteria can convert Asp to Asn, Glu to Gln, and Ser to Cys.

One may often get satisfactory results from the aforementioned methods without having to distract their attentions on a hunt for the appropriate auxotrophs.

18. The method employed here follows a particular strategy, which is:
    a. To prepare a batch of cells conditioned to grow and express protein in ≥80% $D_2O$;
    b. To flash freeze these cells in 1-mL aliquots and store at –80°C, these conditioned cells can then be used at any later date (up to 18 mo) to grow cells and express proteins in ≥ 80% $D_2O$.

    Central to this strategy is the fact that cells already conditioned to grow in $D_2O$ will grow and express proteins in higher levels of $D_2O$ much more reliably than cells not previously conditioned to $D_2O$. The nonconditioned cells may or may not grow at all in $D_2O$. If they do grow, it will take an exasperatingly long time during which they are very likely to lose their plasmids and therefore their ability to express target proteins. In the process of conditioning, which involves multiple growths under varying concentrations of $D_2O$, it is possible for the cell to lose its plasmids. However, future damage, such as wasting of expensive labels and time, can be minimized by simply talking an aliquot of frozen cells and running a small scale test in $H_2O$ or $D_2O$ (it takes much less time in $H_2O$ than in $D_2O$) without using labeled materials to check if the cells can still produce high levels of protein so you know if the cells have maintained their plasmids through the conditioning process. Cells that have not maintained their plasmids must be discarded, but in our experience, conditioned cells that have retained their plasmids will always

reliably produce high levels of target proteins in ≥ 80% $D_2O$ with $^{13}C$, and/or $^{15}N$ labeling. Thus, most of the "risk" has been eliminated from the process so that expensive amounts of $D_2O$ and isotope labeling ingredients need not be wasted.

19. It is better to try two or three flasks of culture at the same time. These should be prewarmed to the proper incubation temperature prior to inoculation.
20. The cells grown up in 20% $D_2O$ culture may grow successfully in 80% $D_2O$ culture. If so, the 45% step can be avoided.
21. This test does not have to be in $D_2O$ or in limited media, because the only object here is to check the frozen conditioned cells still have retained the plasmid required for protein expression. It is possible to have two flasks of cells conditioned side by side, grown from the same cell stock, and frozen at the same time, in which cells from one flask express well while cells from the other show no expression. Thus, this test is critical.
22. For cells grown to $OD_{600nm}$ = 1.0, use 1 mL to inoculate 10 mL. If the cells are only grown to $OD_{600nm}$ = 0.5, use 1.5 mL per 10 mL. There are no constraints here about using a minimal inoculation volume as there is in the case of expressing uniformly labeled proteins in minimal media in $H_2O$, where the volume used for inoculation is usually 1%.
23. Cells grown on limited media using glucose as a sole carbon source are forced to make their amino acids and all other biosynthetic intermediates from the metabolism of glucose, which involves breaking it down into simple molecules (in particular the various intermediates of the tricarboxylic acid and citric acid cycles) and using these to build up the carbon skeleton of all the cell components. In this process, most of the hydrogen in the carbon skeleton is exchanged with environmental hydrogen, which are deuterons in $D_2O$. Also, by bypassing these metabolic processes and using media (such as TY), which involve the simple enzymatic degradation of cellular components and their straightforward assimilation into cellular anabolism, the exchange of skeletal carbon protons with deuterons from the environment will be severely inhibited. Therefore, cells grown in NZ amine or similar enriched media do not make a good starter culture for the preparation of deuterated proteins, not only are they not metabolically suited to do the job, but there might still be a considerate amount of enriched media left unconsumed. This consideration is even more important if the deuterated proteins are also $^{13}C$, $^{15}N$-labeled because the cells will have access to unlabeled carbon and nitrogen sources.
24. With the expression of $^{13}C$-labeled protein, for maximal incorporation of $^{13}C$ into the deuterated proteins being expressed, it is necessary to remove any unlabeled from the overnight culture. This can be accomplished by the following procedures: the cells from the overnight culture should be gently centrifuged out of solution (avoid temperature shock at this stage, i.e., centrifuge at the same temperature as the cells are grown at, with prewarmed, previously autoclaved centrifuge tubes. After centrifuging, pour off the supernatant (save for $D_2O$ recycling), resuspend the pellet gently with a few mL of expression medium containing $^{13}C$

and $^{15}N$ labeled ingredients, then carefully pour this back into the flask containing $^{13}C$ and $^{15}N$ medium in 99.9% $D_2O$.

25. Because of the higher level of $Mg^{2+}$ and $Fe^{2+}$ salts used here and their lower solubility in $D_2O$ than in $H_2O$, visibly high levels of precipitates may form during incubations. These are mainly metal hydroxides. These precipitates have no detrimental effect on cell growth, but they scatter visible light as cells do. Thus, the precipitates may cause a problem in the measurement of cell growth (typically monitored by measuring $OD_{600nm}$). However, as the cells approach mid-log phase, the precipitates clear away when the minerals are consumed by bacteria, $OD_{600nm}$ becomes a reliable indicator for cell growth.
26. Because cell metabolism is slower in $D_2O$, induction will be longer, perhaps 3–4 times longer than in $H_2O$. Accurate induction times should be determined by time-courses.
27. This method is good to recycle $D_2O$ used previously to grow cells and/or $D_2O$ accumulated over the years from various NMR samples. A good rule of thumb is: distill until visibly clear, then distill one more time. This distillate will still contain trace amounts of volatile organic compounds, which are toxic to biological organizms so will not support life. It can therefore be stored indefinitely in the refrigerator or at room temperature.
28. This step should be done immediately before use. Because this $D_2O$ will now support life, it should be stored at refrigeration temperature otherwise mould, alga, and other exotica will grow in it and it will have to be recycled again.
29. The most convenient way to determine the percent deuteration of the $D_2O$ recycled is gravimetrically.

## References

1. Wüthrich, K. (1986) NMR of proteins and nucleic acids. Wiley, New York.
2. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR spectroscopy: Principles and Practice*. Academic, San Diego, California.
3. Gardner, K. H. and Kay, L. E. (1998) The use of $^{2}H$, $^{13}C$, $^{15}N$ multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 357–406.
4. Venters, R. A., Calderone, T. L., Spicer, L. D., and Fierke, C. A. (1991) Uniform $^{13}C$ isotope labeling of proteins with sodium acetate for NMR studies: application to human carbonic anhydrase II. *Biochemistry* **30,** 4491–4494.
5. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185,** 60–89.
6. Gagne, S. M., Li, M. X., McKay, R. T., and Sykes, B. D. (1998) The NMR angle on troponin C. *Biochem. Cell Biol.* **76,** 302–312.
7. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., and Dahlquist, F. W. (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **177,** 44–73.

# 21

# Protein Structure Calculation from NMR Data

**Tapas K. Mal, Stefan Bagby, and Mitsuhiko Ikura**

## 1. Introduction

Until 1984, structural information of biomolecules at atomic resolution could only be determined by X-ray diffraction techniques with protein single crystals ***(1)***. In the mid-1980s, Wüthrich and co-workers demonstrated that nuclear magnetic resonance (NMR) spectroscopy ***(2)*** could be used as a technique for protein structure determination ***(3)***. This permits biomolecular structure determination with comparable accuracy to X-ray diffraction, but in a solution environment that is much closer to the physiological milieu than the single crystals required for protein crystallography. Today, many if not most, NMR measurements with proteins are performed with the ultimate aim of determining their three-dimensional (3D) structure. NMR is not a "microscope with atomic resolution" that directly produces an image of a protein. Rather, NMR yields a wealth of indirect structural information from which the 3D structure can only be elucidated by extensive calculations. The first structure determinations of peptides and proteins in solution ***(4–8)*** were fascinating yet tedious and lengthy struggles because of the lack of established NMR techniques and numerical methods for structure calculation. In the early days of structure calculations from NMR data (NOE-derived distance restraints and $^3J$-derived torsion angle restraints), mostly two types of distance geometry algorithms were used: (1) DISGEO, which operated in distance space and used the metric matrix method to convert distance restraints into Cartesian coordinates ***(9)***, and (2) DISMAN, which operated in torsion angle space and used restrained minimization of a variable target function ***(10)***. Subsequent developments in both NMR methodologies and structure calculation methods have made NMR an indispensable tool for determining biomolecular structures. Here we focus on the most commonly used software packages for 3D structure calculations from

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

NMR data, X-PLOR ***(11)***, ARIA (ambiguous restraints for iterative assignment) ***(12,13)***, and DYANA (dynamical algorithm for NMR applications ***(14,15)***. X-PLOR and ARIA use the distance space, whereas DYANA works in the torsion angle space.

## 2. Materials

Certain pieces of hardware and software are essential for structure calculation from NMR data. Hardware options include SUN (http://www.sun.com), Silicon Graphics Inc. (SGI) (http://www.sgi.com), and Hewlett-Packard (HP) (http://www.hp.com) platforms and PC systems using Linux (http://www.linux.com). The most commonly used software packages are X-PLOR ***(11)*** (http://atb.csb.yale.edu/xplor), ARIA ***(12,13)*** (http://www.cmbl-heidelberg.de/nmr/nilges), and DYANA ***(14,15)*** (http://www.mol.biol.ethz.ch/wuthrich/software/dyana).

## 3. Methods

Structure determination from NMR data can be divided into two steps: (1) collection of structural restraints, and (2) calculation of structures using these restraints. The first step is common to all three structure calculation methods under discussion in this chapter. Two types of restraints are commonly used in structure determination from NMR data:

1. Distance restraints derived from NOE (nuclear Overhauser effect) measurements.
2. Dihedral angle restraints derived from the measurement of vicinal coupling constants ($^3J$).

Collection of structural restraints typically involves the following steps:

1. Assign as many NOE cross peaks from NOESY (nuclear Overhauser enhancement spectroscopy) spectra as possible.
2. Measure cross peak intensities of all assigned NOEs (*see* **Subheading 4.**).
3. Classify NOEs into three different categories, for example "short range" ($d(i, i+1)$), "medium range" ($dij(1<|i-j|\leq 4)$), and "long range" ($dij(|i-j|\geq 5)$).
4. Translate NOE intensities into different classes of upper distance bounds (typically, 1.8–2.8, 1.8–3.5, and 1.8–5.0 Å) (*see* **Subheading 4.**).
5. Measure coupling constants such as $^3J_{HN-H\alpha}$, $^3J_{H\alpha-H\beta}$, $^3J_{H\alpha-N_{i+1}}$ (*see* **Subheading 4.**).
6. Delineate secondary structure elements using NOE, $^3J$, and chemical shift data (*see* **Subheading 4.**).
7. Where possible, assign stereo-specific geometry for diastereo or prochiral groups utilizing both NOE and $^3J$ data (*see* **Subheading 4.**).
8. Assign hydrogen bond restraints from exchange, NOE and $^3J$ data (*see* **Subheading 4.** for details: it is advisable not to use these restraints in preliminary structure calculations but to add them in the structure refinement protocols).

## 3.1. Structure Calculation Methods

### 3.1.1. X-PLOR

NMR data alone are not sufficient to determine the positions of all atoms in a biological macromolecule, but must be supplemented by information about the covalent structure of the protein (the amino acid sequence, bond lengths, bond angles, chiralities, and planar groups) as well as by the steric repulsion between nonbonded atom pairs. Details of structure calculation by X-PLOR are given:

1. Build a protein structure file (psf) from the protein sequence using the standard X-PLOR topology (topallhdg.pro) and peptide bond linkage (toph19.pep). This psf will contain the following information: atom names, types, charges, and masses; residue names and segment names and a list of bond terms, angle terms, dihedral terms, improper terms, explicit hydrogen-bonding terms, explicit nonbonded exclusions, and nonbonded group partitions. It does not contain atomic coordinates, parameters, or restraints.
2. Generate a template structure using standard X-PLOR parameter (parallhdg.pro) and molecular structure (psf) files. The template structure will have an extended conformation with good local geometry and no nonbonded contacts.
3. Create NOE restraints table (*see* **Subheading 4.**).
4. Create dihedral angle restraints table (*see* **Subheading 4.**).
5. Select an appropriate potential function for NOE restraints. The most commonly used potential function is a flat bottom (square well) potential with a soft asymptote ***(11)***.
6. Produce structure from psf with randomized $\phi$ and $\psi$ angles. $\chi^i$-angles are not affected.
7. Energy minimization using Powell gradient function (50 steps) to remove nonbonded interaction.
8. Molecular dynamics-based simulated annealing (MDSA). Typically, MDSA from an extended template structure with randomized $\phi$ and $\psi$ angles works in four stages: a high-temperature search phase, two cooling phases, and a final-energy minimization step. During the high-temperature molecular dynamics (MD) search phase, a low repulsion energy value is used to allow the atoms to pass through each other and to increase the convergence rate. The temperature is reduced from 2000 K to 1000 K in the first cooling phase, and all weights on the different energy terms are gradually brought to their final values (*see* **Table 1**). The second cooling phase comprises a slow cooling from 1000 K to 100 K. The final stage comprises 1000 gradient energy minimization steps using the final weighting values of the various energy terms (*see* **Table 1**).

#### 3.1.1.1. Analysis of Structures Derived from MDSA

1. Restraints violations: Following calculations, the structures must be analyzed to determine whether they fulfil the given experimental restraints. Usually a viola-

**Table 1**
**Simulated Annealing Protocol**

| | Stage | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Temperature[a] (K) | 2000 | 2000 → 1000 | 1000 → 100 | 100 |
| Number of steps | 6500 | 3500 | 3000 | 1000 |
| Parameters and force constants | | | | |
| $k_{NOE}$ (kcal mol$^{-1}$ Å$^{-2}$) | 10 → 50 | 50 | 50 | 50 |
| $k_{repel}$ (kcal mol$^{-1}$ Å$^{-4}$) | 0.002 | 0.01 → 4.0 | 4.0 | 4.0 |
| $k_{dihedral}$ (kcal mol$^{-1}$ rad$^{-2}$) | 5 | 200 | 200 | 200 |
| repel | 0.9 | 0.9 → 0.75 | 0.75 | 0.80 |

[a]The temperature is maintained by coupling to a bath ***(16)*** with a coupling constant of 10 ps$^{-1}$.

tion greater than approx 1 Å is an indication that there is a serious problem, especially if any such violation occurs repeatedly across the ensemble of structures. It is generally useful to identify the region of structures where these restraints are violated and critically inspect them using interactive computer graphics and check the corresponding NOE assignment and volume integration.

2. Atomic root-mean-square deviations (RMSD): The RMSD of the ensemble from the mean of the ensemble is a test that is commonly used to determine the "precision" of the structures, or how close the calculated structures are to each other.
3. Torsion angle distributions: Plots of the $\phi$, $\psi$, $\chi^i$-torsion angles vs amino acid sequence of the protein and Ramachandran plots are useful to analyze local conformations and this allows assessment of the local geometry of the structures.
4. Quality assessment: Programs are available to perform a "quality check" of the structures derived from NMR data, for example, PROCHECK-NMR ***(17)*** and WHATIF ***(18)***. These software packages assess the quality of structures by comparing structural parameters with their values in databases derived from high resolution X-ray structures.

#### 3.1.1.2. Additional Restraints for Structure Refinement

Additional structural restraints can be added to the distance and torsion angle restraints for structure calculation in the final refinement of the structures. For example, $^{13}$C secondary chemical shifts ***(19)***, which are related to backbone $\phi$ and $\psi$ angles, $^{1}$H chemical shifts ***(20,21)***, which are influenced for example by short-range ring-current effects from aromatic groups, magnetic anisotropy of C = O and C–N bonds ***(22)*** and residual dipolar couplings that give informa-

tion on angles between covalent bonds and globally defined axes in the molecule ***(23)***. Although the inclusion of these restraints has little impact on precision, the accuracy of the structures is improved ***(24)***.

### *3.1.2. ARIA*

ARIA has recently been developed to automate the structure calculation from NMR data using X-PLOR following assignments of unambiguous NOEs ***(12,13)***. In general, ARIA is an extended version of X-PLOR. It is a combination of FORTRAN subroutines linked to X-PLOR and scripts that convert data formats and control the overall flow of iterative structure calculations.

The most time-consuming step in NMR structure determination is assignment of NOEs. Often, several protons have the same chemical shift. An NOE crosspeak involving such degenerate protons cannot be converted directly into a distance restraint and used in structure calculation. In normal practice, NOEs that can be assigned unambiguously are used to calculate preliminary 3D structures and then additional NOE cross peaks can be assigned on the basis of these structures ***(25)***. The additional restraints are used to calculate a second generation of structures, which, in turn, is then used to obtain more NOE assignments. This is highly time consuming and laborious. ARIA can automate this process and also identify NOEs that do not provide valid structural information (such as spectral artifacts) and reject them from the structure calculation. A typical ARIA protocol involves the following steps:

1. Assign all possible unambiguous NOEs from NOESY spectra.
2. Determine distance restraints from unambiguous NOEs (*see* **Subheading 3.1.1.**).
3. Make peaklists that contain both unambiguous and ambiguous NOE information.
4. Structure calculation with an extended version of X-PLOR. The following steps are automated within ARIA:
   a. Calculation of a set of structures from unambiguous restraints.
   b. Selection of a subset of lowest energy structures for iterative assignment of ambiguous NOEs.
   c. Calibration of the ambiguous NOEs to distance restraints called ambiguous distance restraints (ADRs) on the basis of the subset of lowest energy structures and selection of the NOEs that make the greatest contribution to the ambiguous NOE cross peaks.
   d. Calculation of another set of structures with unambiguous NOE restraints and assigned ADRs.
   e. Analysis of the violations from the calculated structures and rejection of those NOEs that are constantly violated.
   f. Repetition of **steps 2–5** until completion of the user-defined number of iterations or until no significant changes in structures and data sets are detected.
5. Analysis of the structures obtained from ARIA (*see* **Subheading 3.1.1.1.**).

#### 3.1.2.1. Ambiguous distance restraints

An ambiguous NOE cross peak at the chemical shift coordinates *F1* and *F2* contains contributions from all proton pairs with the same chemical shifts. On the basis of the isolated spin pair approximation (ISPA), an ambiguous NOE can be treated as the sum of the inverse sixth powers of individual proton-proton distances assuming that the proportionality factor is identical for all protons:

$$\mathrm{NOE}_{F1,F2} \propto \sum_{a=1}^{N_\delta} d_a^{-6}$$

where $a$ runs through all $N_\delta$ contributions to a cross peak at frequencies *F1* and *F2*, and $d_a$ is the distance between two protons corresponding to the $a$th contribution. The ambiguous NOE corresponds thus to a "$d^{-6}$ summed distance" $D$:

$$\bar{D} = \left(\sum_{a=1}^{N_\delta} d_a^{-6}\right)^{-1/6}$$

This distance is determined from preliminary structures calculated from unambiguous restraints.

#### 3.1.2.2. Distance Target Functions

During the structure calculation the distances in the structure are typically held to upper and lower bounds according to distance restraints by a gradient-bound flat bottom potential with a soft asymptote that takes care of any large violation ***(26,27)***. The energy of a single distance restraint is

$$E_{\mathrm{NOE}} = k_{\mathrm{NOE}} \begin{cases} (L-D)^2 & \text{if } D < L \\ 0 & \\ (D-U)^2 & \text{if } U < D \le U + \sigma \\ \alpha + \beta(D-U)^{-1} + \gamma(D-U) & \\ & \text{if } D > U + \sigma \end{cases}$$

where $D$ is the distance measured in the current structure model or a $(\sum d_a^{-6})^{-1/6}$ distance, $k_{\mathrm{NOE}}$ is the energy constant, and $U$ and $L$ are the upper and lower bounds of the interproton distances, respectively.

$$U = (d_{\mathrm{ref}}^{-6} \times V/V_{\mathrm{ref}})^{-1/6} + \Delta^+$$

$$L = (d_{\mathrm{ref}}^{-6} \times V/V_{\mathrm{ref}})^{-1/6} + \Delta^-$$

where $\Delta^+$ and $\Delta^-$ account for errors as a result of exchange, motion and spin diffusion; $d_{\mathrm{ref}}$ is the distance from the iteratively calculated structures as $<d^{-6}>^{-1/6}$ averaged over all values for which the distance is smaller than a cutoff (3–6 Å);

$V_{ref}$ is evaluated as the arithmetic average over all corresponding volumes. For a good data set, the best estimated value for $\Delta^+$ is suggested to be max(0.15, 0.15 + 0.08[$D$ – 2.6]) and $\Delta^-$ is 0.15$D$ ***(28)***. The parameter $\sigma$ determines the distance at which the potential switches to asymptotic behavior, $\gamma$ is the asymptotic slope of the potential, the coefficients $\alpha$ and $\beta$ are determined such that $E_{NOE}$ is continuous and differentiable at the point $U+\sigma$. If $D$ is between $L$ and $U$, the energy and gradient are zero. For large restraint violations, the force approaches a maximum value or can be decreased depending on $\alpha$ and $\beta$. This makes the optimization more stable and improves convergence by permitting transient large violations during calculation and thus allows the structures to escape deep local minima. This is important for structure calculation with ADRs.

#### 3.1.2.3. Assignment of ADRs

The criterion used in ARIA is based on the estimation of the relative size of contributions of different assignment possibilities to the peak volume. For each contribution $k$ to the ambiguous NOE, the minimum or average distance ($D^k_{min}$ or $D^k_{av)}$ is determined from the calculated structure. The contribution $C^k$ of the assignment $k$ to the cross peak is estimated as:

$$C^k = (D^k)^{-6} / \sum_i^{N\delta} (D_a^i)^{-6}$$

The $C^k$ are then sorted according to size, and the largest contribution ($Np$) is chosen such that

$$\sum_i^{N\delta} C^i > p$$

The cutoff parameter $p$ can be varied for different iterations, in general starting from values close to 1.0 for the first iteration to 0.8 in the last iteration. The smaller the final value of $p$ chosen, the fewer peaks remain ambiguous.

#### 3.1.2.4. Advantages of ARIA Over General X-PLOR

1. Ambiguous data can be used from the beginning of the structure calculation.
2. Hydrogen bonds are very difficult to assign, especially at the termini of secondary structure elements, and in irregular structures. Hydrogen bond restraints can be used as ambiguous restraints in ARIA.
3. Sometimes it is difficult to know the disulphide bond pattern in a protein. Disulphide bond restraints can be input as ambiguous restraints in ARIA.

### *3.1.3. DYANA*

DYANA ***(14,15)*** calculates solution 3D structures of biomolecules from distance restraints and torsion angle restraints collected from NMR experiments by performing simulated annealing and molecular torsion angle dynamics (TAD) using variable target functions. Both X-PLOR and ARIA also use

molecular dynamics and simulated annealing with variable target functions but work in Cartesian coordinates. The principal differences of TAD from simulated annealing in Cartesian coordinates are:

1. It works with internal coordinates rather than Cartesian coordinates.
2. The number of degrees of freedom in TAD is almost 10 times smaller as the covalent structure parameters such as bond lengths, bond angles, chiralities, and planarities remain fixed at their optimal values during structure calculation.
3. Strong potentials are required to preserve the covalent structure and geometry in conventional Cartesian space molecular dynamics whereas a soft potential function is used in TAD, as the concomitant high frequency motions are absent.
4. TAD gives higher efficiency structure calculation as it uses longer time steps for the numerical integration of motions.

In DYANA, the molecules are treated as a tree structure consisting of a base rigid body that is fixed in space and n rigid bodies, which are connected by rotatable bonds ***(29)***. The degrees of freedom are exclusively torsion angles, i.e., rotation about single bonds. Each rigid body is made up of one or several mass points (atoms) for which the relative positions are invariable. The tree structure starts from a "base," typically at the *N*-terminus of the polypeptide chain, and terminates with "leaves" at the ends of side chains and at the *C*-terminus. The rigid bodies are numbered from 0 to n and the base has the number 0. A typical DYANA protocol involves the use of CALIBA and gridsearch to create a starting conformation with all torsional angles as independent uniformly distributed random variables (discussed later). This is followed by simulated annealing and energy minimization as follows:

1. Perform a short minimization to reduce high energy interactions: 100 conjugate gradient minimization steps are performed at target level 3, i.e., including only distance restraints between atoms up to three residues apart along the sequence, followed by a further 100 minimization steps including all restraints.
2. Exclude all hydrogen atoms from the check for steric overlap, and increase the repulsive core radii of heavy atoms that are covalently bound to hydrogen atoms by 0.15 Å with respect to their standard values. Set the weighting factor for upper and lower distance bounds to 1, for steric lower bounds to 0.5, and for torsion angle constraints to 5 $Å^2$.
3. Perform a TAD calculation at constant high temperature (typically $T_{high} \approx$ 10,000 K). One fifth of all *N* torsion angle dynamic steps are performed at $T_{high}$ (typical value of *N* is 4000 to 8000).
4. Perform the remaining 4N/5 torsion angle dynamic steps with slow cooling to zero.
5. Incorporate all hydrogen atoms to check for steric overlap. Reset the repulsive core radii to their standard values, increase the weighting factor for steric restraints to 2, and perform 100 conjugate gradient minimization steps with inclusion of all restraints.

6. Perform 200 TAD steps at zero reference temperature.
7. Perform 1000 conjugate gradient minimization steps including all restraints.

During the TAD calculation, the list of van der Waals lower distance bounds is updated every 50 steps, or, during minimization, whenever a torsion angle has changed by more than 10° since the last update, or after 100 minimization steps.

DYANA is an integrated program, which includes CALIBA (calibration of NOE intensity vs distance restraints) and a versatile multidimensional grid-search algorithm. CALIBA calibrates NOE intensities into distance restraints. It has different calibration functions for backbone, side-chains and methyl groups. The calibration functions are $V = A/r^6$, $V = B/r^4$, $V = C/r^4$ where $V$ is the peak volume and $r$ is the corresponding distance. The parameters $A$, $B$, and $C$ are either user-defined or calculated automatically (typically, $B = A/d_{\min}^2$ and $C = B/3$, $d$min = 2.4 Å). The multidimensional gridsearch algorithm analyzes the local conformation of an arbitrary molecular fragment of a protein involving the three torsional angles $\phi$, $\psi$, and $\chi^1$ of an amino acid residue, determines the stereospecific assignments of β-protons and generates torsion angle restraints.

## 4. Notes

1. NOEs are the essential NMR data for defining the secondary and tertiary structures of a protein because they permit connection of pairs of hydrogen atoms in amino acid residues that may be far apart in the protein sequence, but close in space (less than about 5 Å apart). The NOE arises from the transfer of magnetization between spins coupled by the dipole–dipole interaction in a molecule undergoing Brownian motion in a liquid ***(30–32)***. The intensity of an NOE, i.e., the volume of the corresponding cross peak in a NOESY spectrum ***(31,33,34)***, is related to the distance r between the two interacting spins by

$$V < r^{-6} > f(\tau_c)$$

where $r^{-6}$ is averaged since the distance $r$ may vary in molecules with inherent flexibility. The remaining dependence of the magnetization transfer on motion enters through the function $f(\tau_c)$ that includes effects of global and internal motions of the molecule.

The NOE is quantified by the volume or intensity of the corresponding cross peak in the NOESY spectrum ***(35)***. Because the linewidths can vary appreciably for different resonances, cross peaks should be quantified by integration over the peak area rather than by measuring peak heights.

2. Measured crosspeak (NOE) volumes are translated to distance ranges. The lower bound is determined from the sum of the van der Waals' radii and the upper bound from the NOE intensity. NOEs are usually translated into upper bounds on interatomic distances rather than precise distance restraints because the presence of internal motions, spin diffusion and, possibly, chemical exchange may affect the intensity of an NOE ***(35)***.

Assuming a rigid body, upper distance bounds ($u$) are calibrated using the equation $V = k/u^6$, where $k$ is a constant that can be determined from known distances, for example the sequential distances $d(H_{\alpha i}, HN_{i+1})$ and $d(HN_i, HN_{i+1})$ in a regular secondary structure element ***(36)*** or by reference to a preliminary structure ***(37)***. The value of $u$ obtained from the above equation may either be used directly as an upper distance bound or NOEs may be calibrated into different classes according to their volume, using the same upper-bound $u$ for all NOEs in a given class. The upper distance bounds are typically put into three classes according to the measured volume of the corresponding NOE cross-peak, for example 2.8 Å (strong), 3.5 Å (medium), and 5.0 Å (weak) ***(7,38)***. This calibration usually yields good results provided that there is a large number of restraints. However, if greater accuracy is required, for example when ligand-binding sites are being studied, a means of obtaining tighter distance restraints from NOE peak intensities is necessary. The full relaxation matrix is commonly used to achieve this ***(39,40)***.

3. Two lines from a typical NOE distance restraint file:

```
assign (resid 1 and name HA)     (resid 2 and name HN)    2.0 0.2 0.8
assign (resid 1 and name HG1#)   (resid 31 and name HA)   2.5 1.3 5.5
```

The first statement selects the atom of residue number 1 and the second statement selects the atom of residue number 2 or 31. The interpretation of the real numbers is dependent on the particular restraining function used for NOE restraints. Here, the first number is deduced from the NOE intensity and the third number is the error value (to account for exchange, spin diffusion, chemical exchange, error in integration of peaks, and so on).

4. NOEs that involve groups of protons with degenerate chemical shifts, in particular methyl groups, are commonly referred to pseudoatoms located at the geometric center of the protons that they represent, and the upper bound is increased by a pseudoatom correction equal to the proton–pseudoatom distance ***(41)***. Programs for automated pseudoatom distance corrections in NOE tables include that written in FORTRAN by M. Nilges (EMBL, Heidelberg) with a C version by M. Osawa (OCI, Toronto) (available from our website — http://nmr.oci.utoronto.ca/ikura/datasoft.html).

In X-PLOR, the setup of pseudoatoms is accomplished by using the NOE assign statement with multiple protons in either atom selection. For example, a medium-range NOE from an Ala methyl group of residue number 1 to the HN proton of the residue number 12 can be written as:

```
assign (resid 1 and name HB#) (resid 12 and name HN) 3.0 1.8 3.1
```

This assign statement sets the lower bound to 1.2 Å and the upper bound to 6.1 Å. The additional pseudoatom correction (1.1 Å) ***(41)*** is added to upper distance bounds. Pseudoatoms (multiple atom selections) should be used primarily for unresolved NOE cross peaks like those of methyl groups, prochiral centers, and aromatic rings. In the case of stereospecific assignments, the distances should be specified explicitly.

**Table 2**
**Pseudoatom Representation for Some Amino Acids Used in the Structure Determination of Proteins from NMR Data**

| Residue | Pseudoatom representation | | $^1$H atoms represented |
|---|---|---|---|
| | X-PLOR/ARIA | DYANA | |
| Gly | HA# | QA | α-methylene |
| Ala | HB# | QB | β-methylene |
| Val | HG1#, HG2# | QG1, QG2 | γ1-, γ2-methyl |
| | HG# | QQG | all six g-methyl |
| Ile | HG1#, HG2# | QG1, QG2 | γ1-methylene, |
| | HD# | QD | γ2-methyl |
| Leu | HD1#, HD2# | QD1, QD2 | δ1-, δ2-methyl |
| | HD# | QD | all six δ-methyl |
| Pro | HB#, HG#, HD# | QB, QG, QD | β-, γ-, δ-methylene |
| Ser, Asp, Cys, His, Trp | HB# | QB | β-methylene |
| Thr | HG# | QG | γ-methylene |
| Asn | HD2# | QD2 | δ2-amido |
| Glu | HB#, HG# | QB, QG | β-, γ-methylene |
| Gln | HB#, HG# | QB, QG | β-, γ-methylene |
| | HE2# | QE2 | ε2-amido |
| Phe, Tyr | HD#, HE# | QD, QE | δ1- and δ2-ring, ε1- and ε2-ring |

The pseudoatom nomenclature used in X-PLOR and ARIA is different from DYANA and is listed in **Table 2**. A universal nomenclature of pseudoatoms for NMR structure calculation and representation has been recommended ***(42)***.

5. Calcium-binding proteins require an additional distance restraints list to account for $Ca^{2+}$–protein interactions. For example, calmodulin binds to four $Ca^{2+}$, which are located in EF-hand loop regions ***(43,44)***. The distance list shown below was used in calculation of a calmodulin–peptide complex structure ***(45)***. In the list, sites I, II, III, and IV indicate the four $Ca^{2+}$-binding sites in calmodulin. Calcium atoms are assigned the residue numbers 149–152.

```
! site I
assign (segid A and resid 20 and name OD2)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 22 and name OD1)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 24 and name OD2)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 26 and name O)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 31 and name OE1)(resid 149 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 31 and name OE2)(resid 149 and name CA2) 2.5 0.8 0.3
! site II
```

```
assign (segid A and resid 56 and name OD2)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 58 and name OD2)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 60 and name OD1)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 62 and name O)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 67 and name OE1)(resid 150 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 67 and name OE2)(resid 150 and name CA2) 2.5 0.8 0.3
! site III
assign (segid A and resid 93 and name OD2)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 95 and name OD2)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 97 and name OD1)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 99 and name O)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 104 and name OE1)(resid 151 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 104 and name OE2)(resid 151 and name CA2) 2.5 0.8 0.3
! site IV
assign (segid A and resid 129 and name OD2)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 131 and name OD2)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 133 and name OD2)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 135 and name O)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 140 and name OE1)(resid 152 and name CA2) 2.5 0.8 0.3
assign (segid A and resid 140 and name OE2)(resid 152 and name CA2) 2.5 0.8 0.3
```

6. The vicinal scalar coupling constant, $^3J$, between atoms separated by three covalent bonds can provide useful geometric information that is complementary to that from the NOE data. In contrast to NOEs, the coupling constants give information only on the local conformation. They are nevertheless important to define accurately the local conformation, to obtain stereo-specific assignments for diastereotopic protons (usually β protons), and to detect torsion angles (usually $\chi^2$) that occur in multiple states. Vicinal scalar coupling constants can be translated into dihedral angles by a Karplus-type equation ***(46)***:

$$^3J(\theta) = A\cos^2\theta + B\cos\theta + C$$

where the parameters $A$, $B$, and $C$ are constants and are determined for various types of coupling constant by a best fit of the measured $^3J$ values to the corresponding values calculated with the above equation from known structures. The most commonly used Karplus relations in proteins are given in **Table 3**.

7. The following is an example of a dihedral angle restraint table.

```
assign   (resid 2 and name C) (resid 3 and name CA)
         (resid 3 and name N)(resid 3 and name C)      1 –120 40.0 2 {* 9 Hz *}
assign   (resid 3 and name C) (resid 4 and name CA)
         (resid 4 and name N)(resid 4 and name C)      1 –120 50.0 2 {* 8 Hz *}
```

The four selections of each assign statement specify the particular dihedral angle. The first number after the selections specifies the energy constant in kcal/mol$^{-1}$rad$^{-2}$, the second number specifies degrees to which the dihedral angle is restrained, the third number specifies the range around the restrained angle, and the last number specifies the exponent of the restraining function ***(11)***.

8. Secondary structures in proteins have characteristic NOE patterns and $^3J$ coupling constants ***(3)***. These two parameters have extensively been used to assign secondary structures (α-helix, $3_{10}$-helix, β-sheet and coil) in proteins, the details

**Table 3**
**The Most Commonly Used Karplus Relations, $^3J(\theta) = A\cos^2\theta + B\cos\theta + C$, for Proteins to Obtain a Torsion Angle θ from the Corresponding $^3J$ Coupling Constant**

| Angle | Coupling | A(Hz) | B(Hz) | Offset C(Hz) | (degree)[a] | Ref. |
|---|---|---|---|---|---|---|
| φ | $H^N$–$H^\alpha$ | 6.98 | –1.38 | 1.72 | –60 | ***47*** |
| | $H^N$–C' | 4.32 | 0.84 | 0.00 | 180 | ***47*** |
| | $H^N$–$C^\beta$ | 3.39 | –0.94 | 0.07 | 60 | ***47*** |
| ψ | $H^\alpha$–$N_{i+1}$ | –0.88 | –0.61 | –0.27 | –120 | ***48*** |
| χ1 | $H^\alpha$–$H^\beta$ | 9.5 | –160 | 1.80 | –120/0 | ***49*** |
| | N–$H^\beta$ | –4.40 | 1.20 | 0.10 | 120/–120 | ***50*** |
| | C'–$H^\beta$ | 7.20 | –2.04 | 0.60 | 0/120 | ***51*** |

[a]Difference between θ and the standard torsion angle φ, ψ, and $\chi^1$.

of which have been documented elsewhere ***(3)***. Recently, the chemical shifts of $C^\alpha$, $C^\beta$, C', and $H^\alpha$ are also being routinely used for identifying local backbone conformation in proteins ***(52,53)***. The $C^\alpha$ and C' nuclei show an upfield shift in β-strand and a downfield shift in helical structures relative to random coil shifts. Both $C^\beta$ and $H^\alpha$ nuclei exhibit the opposite correlation of a downfield shift in β-strands and an upfield shift in helices. Various methods are available for identifying secondary structure elements from the chemical shifts, such as the chemical shift index (CSI) ***(54)***. We routinely employ Metzler's method ***(55)*** that uses a combination of $C^\alpha$ and $C^\beta$ chemical shifts of $i - 1$, $i$, and $i + 1$ residues.

Information about the secondary structure elements in a protein helps with the structural determination process in two ways. First, it allows deduction of dihedral angle restraints based on regular secondary structures. To this end, a new method called TALOS has recently been developed to extract φ and ψ angle restraints by searching a database for chemical shift and sequence homology ***(56)***. Second, hydrogen-bonding restraints may be added to regions assigned to a regular secondary structure, although caution must be taken to ensure that the amide exchange rate data concur with the secondary structure already deduced from the NOEs, $^3J$-coupling constants, and chemical shifts. Any discrepancy implies a distortion of the regular structure or the presence of flexible regions.

## Acknowledgments

This work is supported by a grant from the Medical Research Council of Canada (MRCC) and National Cancer Institute of Canada (NCIC) to M. I. and by OCI/Amgen Fellowship and NCIC to T. K. Mal and M. Ikura are MRCC Scientist and Howard Hughes Medical Institute of International Research Scholars. We thank Tao Yuan, Hong Qian, and Antonio Pineda-Lucena for useful discussions.

## References


1. Drenth, J. (1994) *Principles of Protein X-ray Crystallography*. Springer, New York.
2. Abragam, A. (1961) *Principles of Nuclear Magnetism*. Clarendon, Oxford.
3. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
4. Arseniev, A. S., Kondakov, V. I., Maiorov, V. N., and Bystrov, V. F. and Ovchinnikov, I. A. (1983) Conformation NMR Annalysis of the spatial structure of Butkus eupeus insectotoxin 15A. *Bioeng. Rhim.* **9,** 1667–1689.
5. Braun, W., Bosch, C., Brown, L. R., Gö, N., and Wüthrich, K. (1981) Combined use of proton-proton Overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. *Biochim. Biophys. Acta* **667,** 377–396.
6. Clore, G. M., Brünger, A. T., Karplus, M., and Gronenborn, A. M. (1986) Application of molecular dynamics with interproton distance restraints to three dimensional protein structure determination: a model study of crambin. *J. Mol. Biol.* **191,** 523–551.
7. Williamson, M. P., Havel, T. F., and Wüthrich, K. (1985) Solution conformation of proteinase inhibitor IIa from bull seminal plasma by $^1$H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182,** 295–315.
8. Zuiderweg, E. R. P., Billeter, M., Boelens, R., Scheek, R. M., Wüthrich, K., and Kaptein, R. (1984) Spatial arrangement of the three a helices in a solution structure of E. coli lac repressor DNA-binding domain. *FEBS Lett.* **174,** 243–247.
9. Havel, T. F. and Wüthrich, K. (1984) A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular $^1$H–$^1$H proximities in solution. *Bull. Math. Biol.* **46,** 673–698.
10. Braun, W. and Gö, N. (1985) Calculation of protein conformation by proton-proton distance constraints: a new efficient algorithm. *J. Mol. Biol.* **186,** 611–626.
11. Brünger, A. T. (1992) *X-PLOR, Version 3. 1. A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven, Connecticut.
12. Nilges, M. (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245,** 645–660.
13. Nilges, M., Macias, M., O'Donoghue, S. I., and Oschkinat, H. (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β spectrin. *J. Mol. Biol.* **269,** 408–422.
14. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273,** 283–298.
15. Güntert P. (1998) Structure calculation of biological macromolecules from NMR data. *Quart. Rev. Biophys.* **31,** 145–237.
16. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690.

17. Laskowski, R. A., Ruilmay, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486.
18. Vriend, G. and Sander, C. (1993) Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallogr.* **26,** 47–60.
19. Kuszewski, J, Qin, J., Gronenborn, A. M., and Clore, G. M. (1995) The impact of direct refinement against $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **106,** 92–96.
20. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1995) The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Ser. B* **107,** 293–297.
21. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. (1996) A potential involving multiple proton chemical shift restraints for nonstereospecifically assigned methyl and methylene protons. *J. Magn. Reson. Ser. B* **112,** 79–81.
22. Brüschweiler, R., Liao, X., and Wright, P. E. (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268,** 886–889.
23. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., and Bax, A. (1997) Use of dipolar $^{1}H–^{15}N$ and $^{1}H–^{13}C$ couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4,** 732–738.
24. Clore, G. M. and Gronenborn, A. M. (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16,** 22–34.
25. Güntert, P., Berndt, K. D., and Wüthrich, K. (1993) The program ASNO for computer-supported collection of NOE upper distance restraints as input for protein structure determination. *J. Biol. NMR* **3,** 601–606.
26. Nilges, M., Gronenborn, A. M., Brünger, A. T., and Clore, G. M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2,** 27–38.
27. Hanggi, G. and Braun, W. (1994) Pattern recognition and self-correcting distance geometry calculations applied to myohemerythrin. *FEBS Lett.* **344,** 147–153.
28. Folmer, R. H. A., Nilges, M., Papavoine, C. H. M., Harmsen, B. J. M., Konings, R. N. H., and Hilbers, C. W. (1997) Refined structure, DNA binding studies, and dynamics of the bacteriophage Pf3 encoded single-stranded DNA binding protein. *Biochemistry* **36,** 9120–9135.
29. Abe, H., Braun, W., Noguti, T., and Gö, N. (1984) Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles in proteins. General recurrent equations. *Comput. Chem.* **8,** 239–247.
30. Solomon, I. (1955) Relaxation processes in a system of two spins. *Phys. Rev.* **99,** 559–565.
31. Macura, S. and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by 2D NMR spectroscopy. *Mol. Phys.* **41,** 95–117.
32. Neuhaus, D. and Williamson, M. P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. VCH, New York.

33. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71,** 4546–4553.
34. Kumar, A., Ernst, R. R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95,** 1–6.
35. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *The Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Clarendon, Oxford.
36. Billeter, M., Braun, W., and Wüthrich, K. (1982) Sequential resonance assignments in protein $^1$H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. *J. Mol. Biol.* **155,** 321–346.
37. Güntert, P., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. J., and Wüthrich, K. (1991) Structure determination of the Antp(C39→S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217,** 531–540.
38. Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., and Gronenborn, A. M. (1986) The three-dimensional structure of α-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrainted molecular dynamics. *EMBO J.* **5,** 2729–2735.
39. Borgias, B. A. and James, T. L. (1989) Two-dimensional nuclear Overhauser effect: complete relaxation matrix analysis. *Methods Enzymol.* **176,** 169–183.
40. Bonvin, A. M., Rullmann, J. A., Lamerichs, R. M., Boelens, R., and Kaptein, R. (1993) Ensemble iterative relaxation matrix approach: a new NMR refinement protocol applied to the solution structure of Crambin. *Proteins* **15,** 385–400.
41. Wüthrich, K., Billeter, M., and Braun, W. (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance restraints with nuclear magnetic resonance. *J. Mol. Biol.* **169,** 949–961.
42. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., et al. (1998) Recommendation for the presentation of NMR structures of proteins and nucleic acids. *Pure Appl. Chem.* **70,** 117–142.
43. Babu, Y. S., Sack, J. S., Greenhough, J. J., Bugg, C. E., Means, A. R., and Cook, W. J. (1985) Three-dimensional structure of calmodulin. *Nature* **315,** 37–40.
44. Wylie, D. C. and Vanaman, T. C. (1988) Structure and evolution of the calmodulin family of calcium regulatory protein, in *Calmodulin* (Cohen, P. and Klee, C. B., eds.), ELSEVIER Science Publishers, Amsterdam, pp. 1–15.
45. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256,** 632–638.
46. Karplus, M. (1963) Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85,** 2870–2871.

47. Wang, A. C. and Bax, A. (1996) Determination of the backbone dihedral angles $\phi$ in human ubiquitin from reparametrized empirical Karplus equations. *J. Am. Chem. Soc.* **118,** 2483–2494.
48. Wang, A. C. and Bax, A. (1995) Reparametrization of the Karplus relation for $^3J(H_\alpha\text{-}N)$ in peptides from uniformly $^{13}C/^{15}N$ enriched human ubiquitin. *J. Am. Chem. Soc.* **117,** 1810–1813.
49. De Marco, A. C., Llinas, M., and Wüthrich, K. (1978) Analysis of the $^1$H-NMR spectra of ferrichrome peptides. I. The non-amide protons. *Biopolymers* **17,** 617–636.
50. De Marco, A. C., Llinas, M., and Wüthrich, K. (1978) $^1H–^{15}N$ spin-spin couplings in alumichrome. *Biopolymers* **17,** 2727–2742.
51. Fischman, A. J., Live, D. H., Wyssbrod, H. R., Agosta, W. C., and Cowburn, D. (1980) Torsion angles in the cystine bridge of oxytocin in aqueous solution. Measurements of circumjacent vicinal couplings between $^1$H, $^{13}$C, and $^{15}$N. *J. Am. Chem. Soc.* **102,** 2533–2539.
52. Wishart, D. S. and Nip, A. M. (1998) Protein chemical shift analysis: a practical guide. *Biochem. Cell Biol.* **76,** 153–163.
53. Venters, R. A., FarmerII, B. T., Fierke, C. A., and Spicer, L. D. (1996) Characterizing the use of perdeuteration in NMR studies of large proteins: $^{13}$C, $^{15}$N and $^1$H assignments of human carbonic anhydrase II. *J. Mol. Biol.* **264,** 1101–1116.
54. Wishart, D. S. and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239,** 363–392.
55. Metzler, W. J., Constantine, K. L., Friedrichs, M. S., Bell, A. J., and Ernst, E. G. (1993) Characterization of the three-dimensional solution structure of human profilin: $^1$H, $^{13}$C and $^{15}$N assignments and global folding pattern. *Biochemistry* **32,** 13,818–13,829.
56. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biolmol. NMR* **13,** 289–302.

# 22

# Shape and Dynamics of a Calcium-Binding Protein Investigated by Nitrogen-15 NMR Relaxation

**Jörn M. Werner, Iain D. Campbell, and A. Kristina Downing**

## 1. Introduction

For a growing number of proteins, structural biology has enabled function to be understood at the atomic level. It is clear that flexibility plays an essential role in determining the biological properties of macromolecules. Nuclear magnetic resonance (NMR) is unique in its ability to probe dynamics of the backbone and side chains on the ms to ps time scale (reviewed in **ref.** ***1***). This powerful method has been exploited in the study of a wide variety of systems, including several calcium-binding proteins ***(2–6)***.

We have used $^{15}N$ relaxation to study the shape and dynamics of a pair of calcium-binding (cb) epidermal growth factor-like domains (EGFs) from human fibrillin-1 in the presence of calcium ***(7)***. Human fibrillin-1 is a large (approx 350 kDa) extracellular matrix glycoprotein, that is localized to 10–12 nm microfibrils. Mutations to this protein have been associated with the Marfan syndrome, an autosomal dominantly inherited disease of connective tissues that is estimated to affect approx 1/5000 in the population, and related disorders ***(8)***. This protein is mainly comprised of multiple tandem repeats of cbEGF domains, and electron microscopy has been used to show that calcium plays a key role in fibrillin-1 and microfibril architecture ***(9,10)***. A large number of mutations have been identified in Marfan syndrome patients, and many of these mutations change residues that are directly involved in calcium binding ***(11)***. Because these mutations are not predicted to disrupt folding or protein–protein interactions, it has been suggested that the amino acid changes may alter the flexible properties of the molecule, leading to increased proteolytic susceptibility or altered biomechanical properties of the microfibril assembly ***(9,12)***.

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

Our research has shown that calcium plays an essential role in stabilizing the cbEGF domain linkage ***(7)***.

In this chapter, we introduce the interpretation of $^{15}N$ relaxation data with a brief overview of the theory. Relaxation of a $^{15}N$ nucleus in the presence of a proton is dominated by the $^{1}H$–$^{15}N$ dipolar coupling and by chemical shift anisotropy (CSA). Ignoring cross-correlation between $^{1}H$–$^{15}N$ dipolar and CSA relaxation, the longitudinal ($T_1$) and transverse ($T_2$) relaxation time constants as well as the [$^{1}H$]–$^{15}N$ heteronuclear NOE are given by the following equations ***(13)***:

$$1/T_1 = d^2/4\,[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2\,J(\omega_N) \quad (1)$$

$$1/T_2 = d^2/8\,[J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + c^2/6\,[4J(0) + 3J(\omega_N)] + R_{ex} \quad (2)$$

$$\mathrm{NOE} = 1 + d^2/4T_1(\gamma_H/\gamma_N)[6J(\omega_H + \omega_N) - \mathrm{J}(\omega_H + \omega_N)] \quad (3)$$

in which $d = (\mu_o h \gamma_N \gamma_H) / (8\pi 2)\langle r_{NH}^{-3}\rangle$ and $c = (\omega_N \Delta\sigma) / (\sqrt{3})$; $\mu_o$ is the permeability of free space; $h$ is Planck's constant; $\gamma_H$ and $\gamma_N$ are the gyromagnetic ratios of the $^{1}H$ and $^{15}N$ spins; respectively; $r_{NH}$ is the N–H bond length; $\omega_H$ and $\omega_N$ are the Larmor frequencies of the $^{1}H$ and $^{15}N$ spins, respectively; and $\Delta\sigma$ is the chemical shift anisotropy of the $^{15}N$ nucleus, assuming axial symmetry and colinearity of the symmetry axis and the N–H bond vector.

Motions on a μs to ms time-scale that can contribute to the transverse relaxation time constant, $T_2$, are modeled as an exchange term, $R_{ex}$, in **Eq. 2**. The quadratic dependence of exchange line broadening with either $B_1$, the field in the rotating frame ***(14–16)***, or with $B_0$, the static magnetic field, has been used to measure $R_{ex}$ ***(17–19)***. Alternatively, exchange can be determined from the ratio of transverse and longitudinal cross-relaxation rate constants resulting from $^{1}H$–$^{15}N$ dipole and $^{15}N$ CSA relaxation interference ***(20)***. In this study, $R_{ex}$ was obtained from measurements of relaxation time-constants at multiple fields. Provided that the spectral density at the proton frequency is small compared to the spectral density at zero frequency, it can been shown that $1/T_2 - 1/(2T_1)$ is a linear function of the square of the spectrometer field strength ***(19)***.

$$1/T_2 - 1/(2T_1) = d^2 / 2J(0) + [2/9\gamma^2_N\,\Delta\sigma^2\,J(0) + A]\,B_0^2 \quad (4)$$

with $y$-intercept $d^2/2J(0)$ and a slope that depends on the value of the chemical shift anisotropy and exchange contributions, $R_{ex} = AB_0^2$. The advantage of this method is that it is independent of the estimation of overall rotational diffusion properties. The precision of the $R_{ex}$ terms is primarily limited by the precision of the determination of the relaxation rate constants and the range of available spectrometer fields. In studies of calcium-binding proteins, it is important to ensure full saturation of the binding sites as most of the effects of calcium exchange on resonance linewidth are then removed.

Anisotropic diffusion models ***(21)*** in combination with the model-free approach first described by Lipari and Szabo ***(22)*** have been successfully applied to describe the overall and internal dynamics of proteins. The spectral density of an axially symmetric diffusion tensor ($D_\perp = D_{xx} = D_{yy}$ and $D_\parallel = D_{zz}$, where $D_{xx}$, $D_{yy}$, and $D_{zz}$ are the principal moments of the diffusion tensor with trace $D$) with internal dynamics is given by

$$J(\omega) = 2/5(S^2 \sum_{i=1}^{3} [A_i\tau_i/1+(\omega\tau_i)^2] + (1 - S^2) [\tau_c/1+(\omega\tau_c)^2] \quad (5)$$

where $A_1 = (1.5 \cos^2 \alpha - 0.5)^2$, $A_2 = 3\sin^2 \alpha\cos^2 \alpha$, $A_3 = 0.75\sin^4 \alpha$, with $\alpha$ being the angle of the N–H bond vector with $D_\parallel$; the correlation times $\tau_1 = (6D_\perp)^{-1}$, $\tau_2 = (D_\parallel + 5D_\perp)^{-1}$, $\tau_3 = (4D_\parallel + 2D_\perp)^{-1}$, and $\tau_c^{-1} = 6D + \tau_e^{-1}$; and $S^2$ is the generalized order parameter (*see* **Fig. 1**). For a spherical diffusion tensor ($D_{xx} = D_{yy} = D_{zz}$) the first term in **Eq. 5** reduces to a single (isotropic) correlation time, $\tau_m = \tau_1 = \tau_2 = \tau_3 = (6D)^{-1}$ with $A_1+A_2+A_3 = 1$. Fast motions on two time-scales are included in the limit that the faster internal correlation time can be neglected ***(23)***. Then the faster motion is characterized by its order parameter $S^2_f$ and the slower motion is characterised by $S^2_s$ and its correlation time $\tau_e$. The spectral density for this model is obtained by exchanging $S^2$ with $S^2_s$ and multiplying the right-hand side of **Eq. 5** with $S^2_f$. The reported order parameter for this model is the product of the order parameter of the faster and slower motion $S^2 = S^2_s S^2_f$.

The dependence of the spectral density function, $J(\omega)$, on the overall diffusion tensor of the molecule provides information on its shape (*see* **Fig. 1**, *see* **Subheading 3.3.**), and the dependence of $J(\omega)$ on the order parameters $S^2$ and the correlation times $\tau_e$ is used to characterise amplitudes and correlation times of the internal dynamics of the protein (*see* **Fig. 1**, *see* **Subheading 3.4.**).

We describe how $^{15}$N-relaxation studies may be used to obtain a description of shape and dynamics of proteins. As an example, we use the study of the thirty-second and thirty-third cbEGF domains from human fibrillin-1 (cbEGF32–33) in the presence of calcium ***(7)***

## 2. Materials

1. A purified sample of a $^{15}$N isotopically enriched protein, typically containing approx 0.2–1.0 m*M* of protein is dissolved in 90%/10% $H_2O/^2H_2O$. In this study, initial samples were prepared containing approx 5 m*M* cbEGF32–33, pH 6.5 in the presence of 12 m*M* $CaCl_2$. To ensure that the data were not affected by sample aggregation, measurements were repeated on final samples containing 2 m*M* cbEGF32-33, 5 m*M* Tris-HCl, pH 6.5, and 12 m*M* $CaCl_2$ (*see* **Note 1**).
2. Two-dimensional [$^1$H]–$^{15}$N correlated spectra were recorded on the protein sample using high-field NMR spectrometers (≥400 MHz). (In some cases, it is desirable to make measurements at multiple fields — *see* **Subheading 3.2., step 9**).

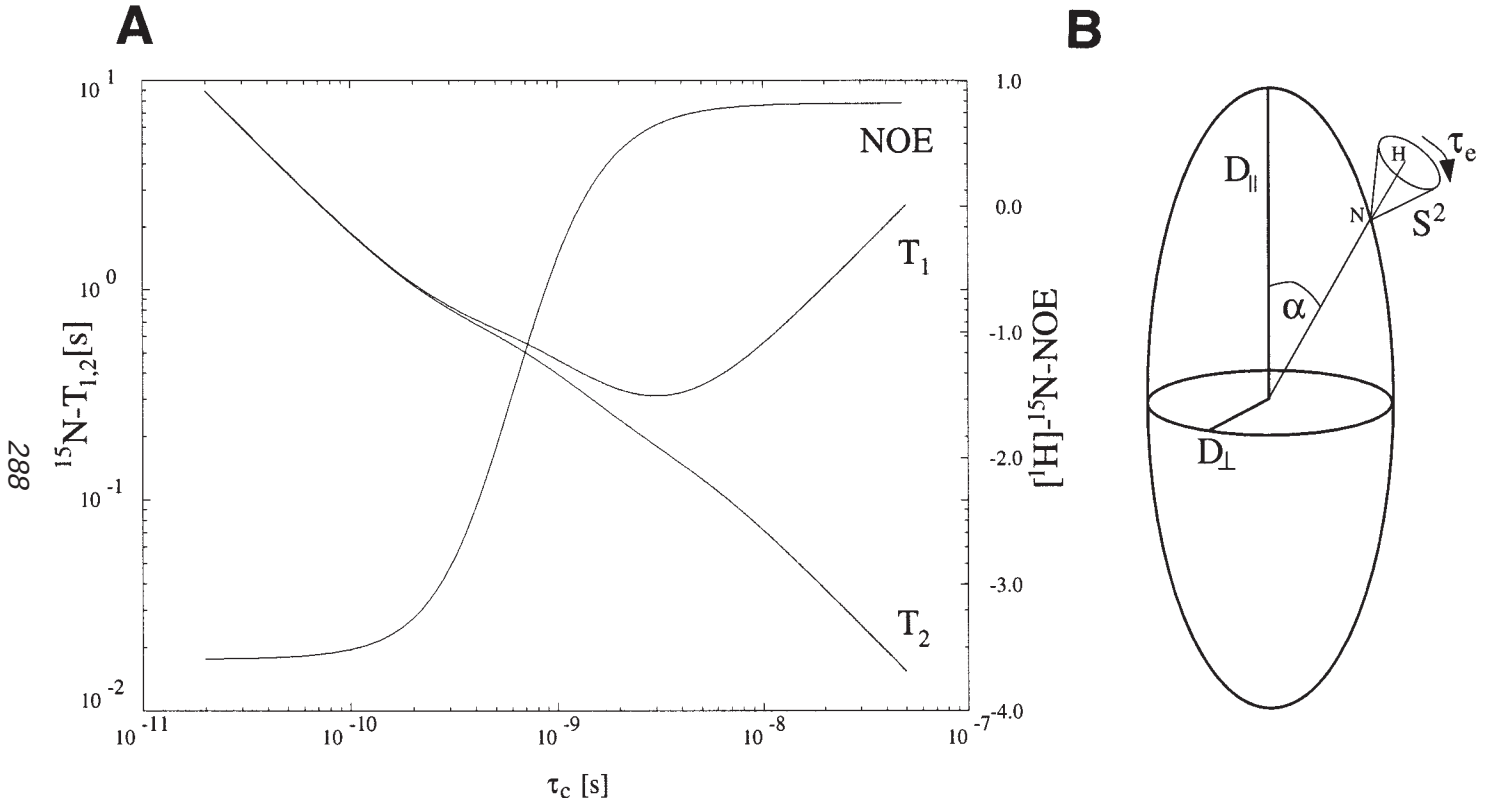


Fig. 1. **(A)** $T_1$, $T_2$ and the heteronuclear NOE as a function of correlation time, $\tau_c$. **(B)** Schematic of a symmetric top with principal axes $D_{\|}$ and $D_{\perp}$. The angle of the N–H bond vector with $D_{\|}$ is labeled $\alpha$. The order parameter $S^2$ and internal correlation time $\tau_e$ are visualized in a diffusion-in-a-cone interpretation.

3. A high-speed computer, with suitable operating system, together with C and FORTRAN compilers is required to run various software packages; these include NMR spectral analysis software such as Felix 2.3 (MSI, Inc., San Diego, CA) or NMRView ***(24)***, molecular visualization software such as MOLMOL ***(25)*** and relaxation data analysis software such as Modelfree4 ***(26,27)***.

## 3. Dynamics by NMR

### *3.1. Data Acquisition*

1. Ideally, all of the data should be collected using one sample, in one experimental session (*see* **Note 2**). It is desirable to preequilibrate the spectrometer and sample at the required temperature. If analysis of the $T_1/T_2$ ratios indicates the presence of unspecific aggregation (*see* **Subheading 3.2., step 8**), then $T_1$ and $T_2$ experiments should be repeated at a lower sample concentration.
2. Usually the $T_1$ and $T_2$ experiments are collected as a series of [$^1$H]–$^{15}$N autocorrelation spectra incorporating preparation periods for the spin states of interest followed by delays during which relaxation occurs (*see* **Note 3**) ***(28,29)***. Transverse relaxation time-constants ($T_2$) are typically measured using a spin-echo sequence (CPMG) with a refocusing interval of less than approx 1 ms between $^{15}$N pulses to prevent the evolution of antiphase magnetization ***(30)***. Dipolar and chemical shift anisotropy cross-correlation are removed by application of proton 180° pulses every few ms ($T_1$) and in the middle of the basic CPMG block ($T_2$) ***(31,32)***. The delays are chosen to sample the intensity decay for up to about 1.5 times the maximal relaxation time-constant. This decay is monitored by following the intensity of the NH peak intensities in the first block of the [$^1$H]–$^{15}$N autocorrelation spectrum at increasing delay times. Typically, a series consists of 8 to 12 spectra and at least one of the spectra should be recorded in duplicate (*see* **Note 4**). The duplicate spectrum is used for estimating peak-height uncertainties.
3. Contributions to the linewidth from motions on the µs-ms time-scale should be investigated by additional experiments (*see* **Note 5**).
4. Inaccuracies in $T_2$ measurements, associated with sample heating, can be reduced by applying a train of $^{15}$N refocusing pulses and delays preceding the usual pulse sequence, such that the total number of $^{15}$N refocusing pulses is the same in each $T_2$ experiment ***(33)***. Heating can be assessed by comparison of the NH chemical shift changes of $T_2$ spectra with the minimum and maximum number of $^{15}$N refocusing pulses.
5. A pair of $^1$H detected [$^1$H]–$^{15}$N heteronuclear NOE experiments ***(28,34)***, with and without NOE, is recorded with the same acquisition times as the $T_1$ and $T_2$ experiments. $^1$H saturation in the NOE experiment is brought about by a train of 120° flip-angle pulses at 10-ms intervals or by a broad-band decoupling field, applied for at least five times the $^{15}$N-$T_1$ .
6. Water saturation is best avoided, especially in the [$^1$H]–$^{15}$N heteronuclear NOE experiments, and the use of water flip-back pulses is recommended for all experiments ***(35)***. Provided that pulsed field gradients are available, coherence selec-

tion, using gradient echoes, yields improved signal-to-noise ratio and good water suppression ***(36)***.

### 3.2. Data Processing and Analysis

1. All spectra recorded in one series are processed identically. Linear prediction (in the $^{15}N$ dimension) and resolution enhancement (in the $^{1}H$ and $^{15}N$ dimensions) can be used to reduce overlap of adjacent peaks (*see* **Note 6**) ***(37,38)***. In the cbEGF32–33 pair, the data were zero-filled to obtain a digitization of less than about 4 Hz (in $F_2$: $^{1}H$) and about 2 Hz per point (in $F_1$: $^{15}N$).
2. The nonoverlapping peaks in the most intense [$^{1}H$]–$^{15}N$ autocorrelation spectrum are picked and assigned and the intensities of these peaks are extracted from each spectrum (*see* **Note 7**).
3. Peak height uncertainties are estimated, either from the baseline noise or from the standard deviation of the difference in peak intensities of two spectra recorded with the same relaxation delay (*see* **Note 8**).
4. The series of [$^{1}H$]–$^{15}N$ correlation peak intensities from the respective experiments are used to obtain the relaxation time-constants of each residue by $\chi^2$ minimization of monoexponential decay functions with initial intensity $I_o$ and decay constants $T_1$ or $T_2$ as free parameters

$$I(t) = I_o e^{-t/T_{1,2}} \tag{6}$$

   Monte Carlo simulations are performed to estimate the uncertainties of the free parameters. The model is accepted if $\chi^2$ of the best fit is within a chosen (typically the ninety-fifth) percentile of the $\chi^2$ distribution of the simulated data (*see* **Notes 7** and **9**).
5. In the heteronuclear NOE experiments, errors of the peak intensities in spectra with and without NOE are estimated from baseline noise for the respective spectrum. In the presence of multiple pairs of heteronuclear NOE spectra averages and standard deviations can be taken over the intensity ratios.
6. The NOE ratio is calculated as the ratio of the intensities in the spectrum with saturation, $I_s$, and without saturation, $I_{ns}$, NOE = $I_s/I_{ns}$, and the uncertainty is obtained by error propagation:

$$\sigma\text{NOE} = I_s / I_{ns} \sqrt{(\sigma I_s / I_s)^2 + (\sigma I_{ns} / I_{ns})^2}.$$

7. The rotational correlation time of the entire molecule is obtained from the average of the $T_1/T_2$ ratios ***(28)***. In case substantial numbers of residues are affected by exchange or fast motion, the average over a subset of residues representative of overall diffusion is used (*see* also **Subheading 3.3., step 2**).
8. A plot of the $T_1$ vs the $T_2$ value of each residue overlaid on parametric curves of $T_1$ and $T_2$ as functions of correlation time $\tau_c$ and order parameter $S^2$ is a convenient way of obtaining a qualitative assessment of the data (*see* **Fig. 2**). Residues whose $T_1$ and $T_2$ values fall on a straight line through the origin, with slope $T_1/T_2$ can be described by a single correlation time. Residues with $T_1$ and $T_2$ values that are shifted to the right (i.e., small $T_2$ values) outside the theoretical curves are

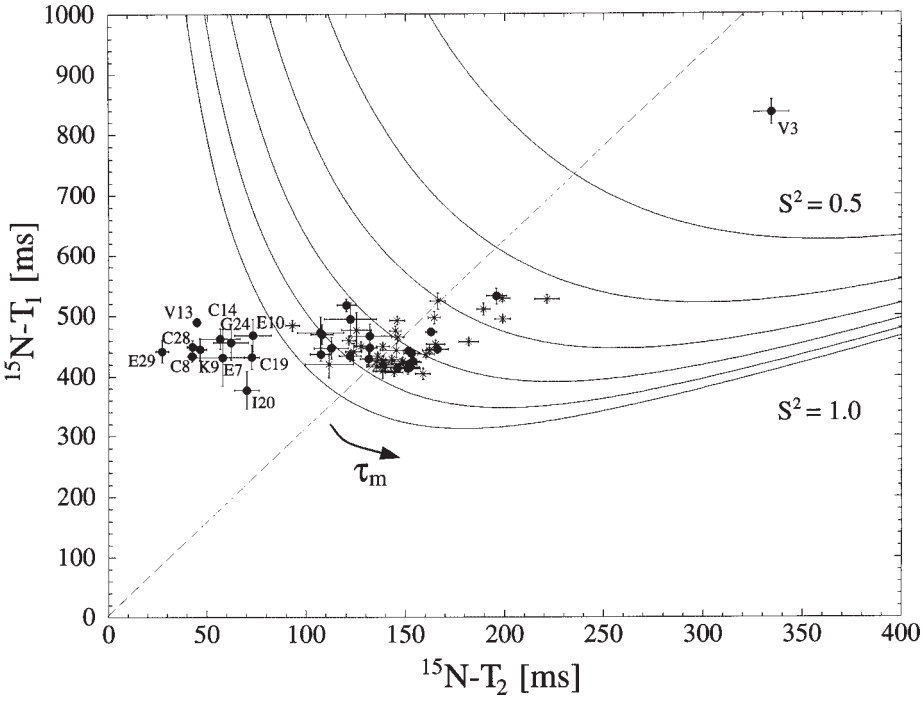


Fig. 2. Backbone $^{15}$N relaxation time-constants for cbEGF32–33 measured at 11.7 *T*. $T_2$ and $T_1$ data of cbEGF32 and cbEGF33 in cbEGF32–33 are shown as circles and stars, respectively and outlying residues are labeled by residue name. The continuous lines are calculated $T_1$ and $T_2$ values as a function of isotropic correlation time $\tau_m$ and the order parameter $S^2$ using the isotropic Lipari and Szabo model (*see* **Eq. 5**). Curves with values of $S^2$ equalling 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 are shown with the highest and lowest values labeled accordingly. The arrow indicates the dependence of calculated $T_1$ and $T_2$ values on a decreasing correlation time.

affected by motions on the μs to ms time-scale ($R_{ex}$). Finding a large proportion of residues outside the theoretical $T_1$ and $T_2$ curves indicates nonideal effects, possibly because of unspecific aggregation. Residues with increased $T_1$ and $T_2$ values, thus shifted to the upper right-hand corner, are subject to motions on a time-scale that is smaller than the overall tumbling time of the molecule. [$^1$H]–$^{15}$N heteronuclear NOEs of those residues are expected to be significantly smaller than the average.

9. Exchange contributions are estimated from model-free analysis (*see* **Subheading 3.4., step 1**) or using the quadratic field dependence of $1/T_2 - 1/(2T_1)$ on $B_0$ (*see* **Note 5**). The slope depends on the value of the chemical shift anisotropy (CSA) and on exchange contributions, $R_{ex}$ (*see* **Note 10**).

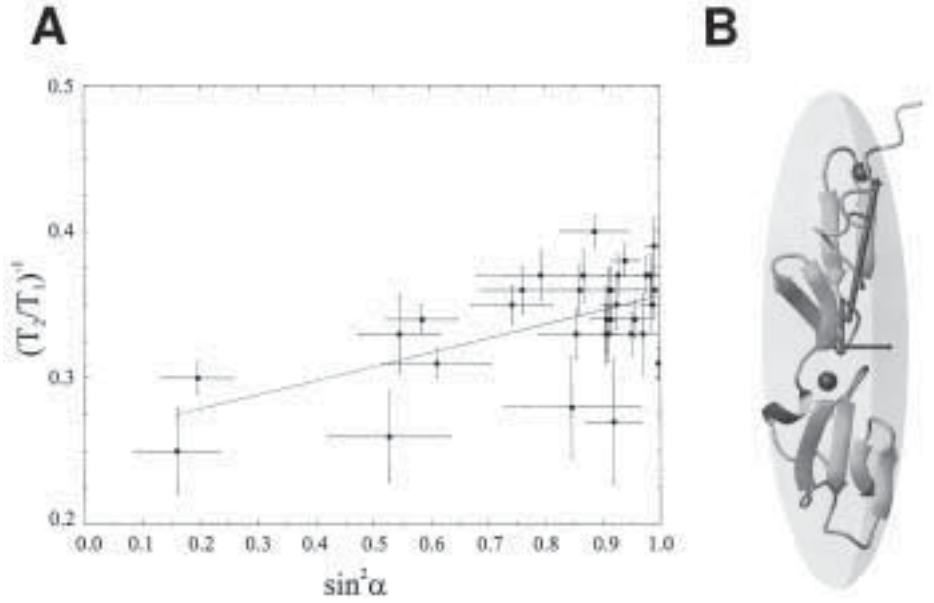


Fig. 3. **(A)** Rotational diffusion anisotropy of cbEGF32-33. The diffusion tensor of each of the 22 structures in the family was calculated and the angle α between the N–H bond vector and the largest principal component of the diffusion tensor was determined. $(T_1/T_2)^{-1}$ ratios are plotted as a function of the averages and standard deviations of the angles α. The linear least-squares fit to the data is shown for illustration of **Eq. 10**. **(B)** Orientation of the unique axis of the diffusion tensor $D_{\|}$ in the molecular frame of cbEGF32-33 depicted using an arrow and a ribbon diagram of the average structure. The coordinate system of the inertia tensor and approximate shape of the molecule are provided as references. The figure was produced using the program MOLMOL ***(25)***.

10. Quantitative data analysis employs similar statistical reasoning at three stages: in derivation of the relaxation time-constants and their associated uncertainties, in determination of a global diffusion model, using a given a set of $T_1$ and $T_2$ values, and in selection of models of internal dynamics for each residue.

    In each case, parameters, $a_1 \ldots a_M$, are obtained by a least-squares fit of the model data, $y(x_i;a_1 \ldots a_M)$, to the experimental data $y_i$ with errors $\sigma_i$ by minimizing $\chi^2$

$$\chi^2 = \sum_i^N [y_i - y(x_i;a_1 \ldots a_M) / \sigma_i]^2 \qquad (7)$$

    The uncertainties of the parameters are determined by Monte Carlo simulations of the fitted values, using the experimental errors. Concomitantly a $\chi^2$ distribution is generated and a fit is deemed adequate, if the $\chi^2$ of the best fit is within a chosen percentile (usually the ninety-fifth percentile) of the distribution ***(39)***.

**Table 1**
**Diffusion Models for cbEGF32-33 at pH 6.5 and T = 35°**

| residue | $D_{\parallel}/D_{\perp}$ | θ(deg) | ϕ(deg) | $D_{aniso}^{(a)}$ | $\chi_{aniso}^{2(b)}$ | $D_{iso}^{(a)}$ | $\chi_{iso}^{2(b)}$ | $Q^{(c)}$ |
|---|---|---|---|---|---|---|---|---|
| core | 1.45±0.07 | 13.2±1.0 | −43.4±3.8 | 3.21±0.48 | 148 | 3.28 | 208 | $3.9e^{-3}$ |

[a] Values are given in $10^{-7}$ Hz.
[b] $\chi^2$ of all fitted residues.
[c] $Q$ is calculated for the comparison of the isotropic and symmetric top models.

Models of increasing complexity (i.e., incorporating increasing numbers of parameters) are selected using a hierarchical approach based on *F*-tests ***(27)***. Models that are more complex are tested only if simpler models are not adequate. Provided that the more complex model is found adequate, the *F*-test is used to determine the statistical significance of the improvement. The *F* of an *n*-parameter model and an *m*-parameter model with χ-squares of $\chi_1^2$ and $\chi_2^2$, respectively is given by:

$$F = \nu_2(\chi_1^2 - \chi_2^2) / (\nu_1 - \nu_2)\chi_2^2 \quad (8)$$

in which $\nu_1$ and $\nu_2$ denote the degrees of freedom of a model calculated as $M - n$, where $M$ is the number of measured values and $n$ the number of free parameters in the model. The significance of an improvement of the $\chi^2$ value for two different models is tested by the probability $Q(\nu_1,\nu_2,F)$ of obtaining a value of $F$ as large as the given $F$ by chance ***(39)***. Hence, small $Q$ values indicate that the more complex model is justified (*see* **Note 11**).

### *3.3. Estimation of Diffusion Tensors*

1. The quantitative description of the overall diffusion tensor is based on the orientations of the N–H bond vectors within a molecular frame (*see* **Note 12**). The orientations are usually supplied by an atomic coordinate (PDB) file.
2. A combination of structure- and dynamics-based criteria are used to establish a subset of residues that are representative of the overall diffusion (*see* **Note 13**). The initial set contains residues in secondary structural elements. This is further refined by excluding residues with low [$^1$H]–$^{15}$N heteronuclear NOEs or $T_2$ values that indicate sub-nanosecond or μs to ms time-scale motion, respectively. In the case of families of NMR-derived structures, the set of selected residues can be further restricted to residues with high backbone torsion angle (ϕ,ψ) order parameters. In the study of cbEGF32-33, a set of 33 residues from the central region of the module pair (termed core in **Table 1**), between Phe32 and Cys69, were chosen as representative of overall diffusion.
3. Isotropic ($D = D_{xx} = D_{yy} = D_{zz}$) and anisotropic ($D_{\perp} = D_{xx} = D_{yy}$ and $D_{\parallel} = D_{zz}$) diffusion models are fitted to the experimental $T_1/T_2$ ratios, $(T_1/T_2)_i^{exp}$, of the selected subset (consisting of $N$ residues) and a given set of atomic coordinates. The analysis can be performed using the program Modelfree4 (available from A.

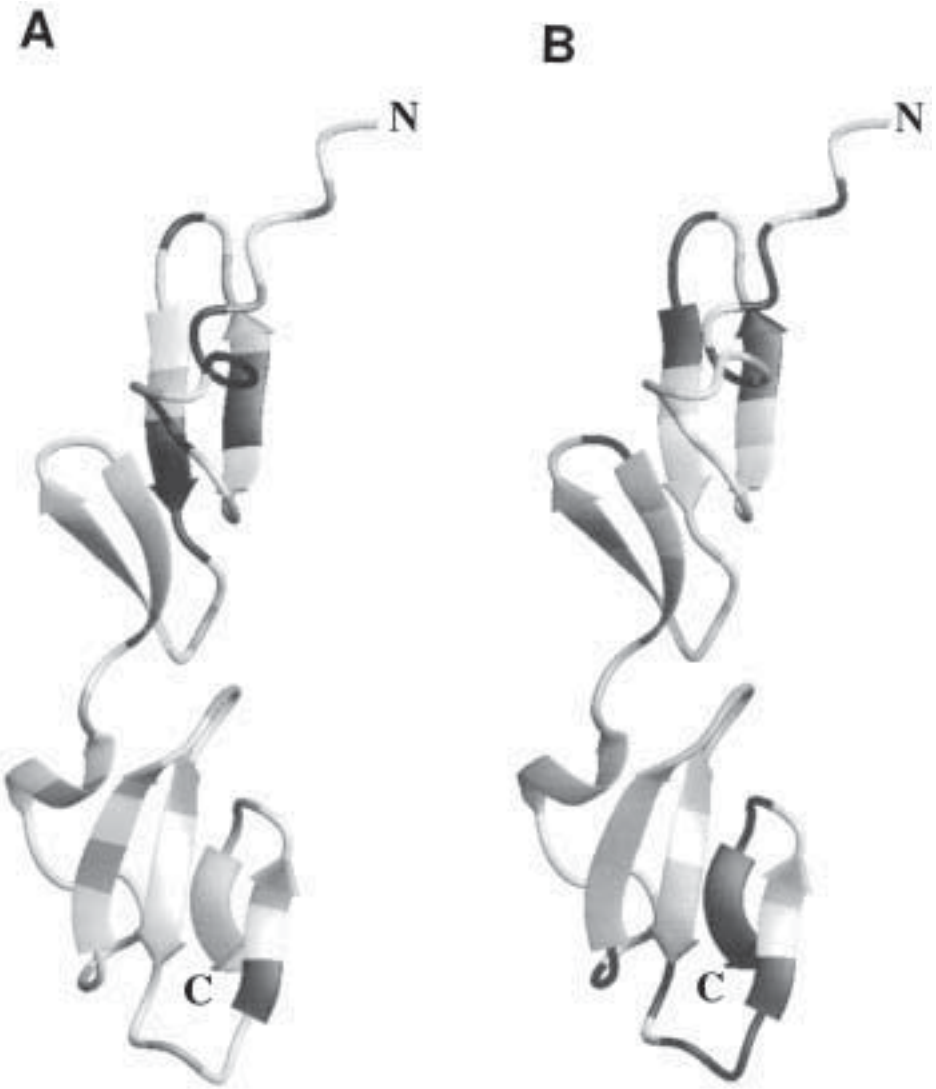


Fig. 4. **(A)** Line-broadening contributions (in $S^{-1}$ at 11.7 T) and **(B)** correlation times for fast motions ($\tau_e$, in ps) from the Modelfree4 analysis of cbEGF32–33 mapped onto a cartoon representation of the structure. In **(A)** $R_{ex}$ values of 0–3 Hz and >3.0 Hz are shown in medium and dark *gray*, respectively. In **(B)** $\tau_e$ values 0–70 ps and >70 ps are shown in medium and dark *gray*, respectively. Large exchange terms are localized principally to the *N*-terminal cbEGF32 domain. Fast time scale motions affect the full length of the construct, and in particular the *N*- and *C*-termini. This figure was produced using the program MOLMOL ***(25)***.

G. Palmer, III, Columbia University). The statistical significance of changes in $\chi^2$ are evaluated as outlined in **Subheading 3.2., step10**, where

$$\chi^2_{tot} = \sum_i^N \left( (T_1/T_2)_i^{exp} - (T_1/T_2)_i^{th} / \sigma(T_1/T_2)_i^{exp} \right)^2 \quad (9)$$

and $(T_1/T_2)_i^{th}$ is a function of the diffusion tensor via **Eqs. 1–3** and **5**, with $S^2 = 1$. For the isotropic model, the isotropic diffusion constant $D$ and for the axially symmetric diffusion model, the isotropic diffusion constant $D$, the axial ratio of the diffusion tensor, $D_{||}/D_{\perp}$, and its orientation in the molecular frame $\theta$ and $\varphi$ are obtained by minimization of **Eq. 9**. Averages of these quantities for a family of 22 NMR structures of cbEGF32-33 ***(40)*** are shown in **Table 1** and **Fig. 3A**.

4. The $\chi^2$ values of the isotropic fit $\chi^2_{iso}$ and the axially symmetric fit $\chi^2_{axial}$ are used to calculate the probability $Q$ that the improvement in $\chi^2_{axial}$ is obtained by chance (*see* **Subheading 3.2., step 10** and **Table 1**). The degrees of freedom are $N$–1 for the isotropic model and $N$–4 for an axially symmetric tensor.
5. For an axially symmetric diffusion tensor the dependence of the $(T_1/T_2)^{-1}$ ratios $R$, as a function of the angle $\alpha$ of the N–H bond vector with the unique axis of the diffusion tensor can be approximated by ***(41)***:

$$R(\alpha) = R(0)(1 + \varepsilon \sin^2 \alpha) \quad (10)$$

with $\varepsilon = (D_{||}/D_{\perp}-1)$. Provided that the analysis indicates some anisotropic motion, a plot of the $(T_1/T_2)^{-1}$ as a function of $\sin^2\alpha$, yields a straight line with $y$-intercept $R(0)$ and slope $R(0)(D_{||}/D_{\perp}-1)$ (*see* **Fig. 3**). Sampling of a wide range of angles is important for obtaining a robust fit and a reliable estimate of the anisotropy.

### 3.4. Model-Free Analysis of $T_1$, $T_2$ and $^1H$–$^{15}N$ NOE Data

1. Internal dynamics is characterized using the model-free approach of Lipari and Szabo ***(22)*** as implemented in Modelfree4 ***(26,27)***. Parameters of internal dynamics are obtained by $\chi^2$ minimization (*see* **Subheading 3.2., step 10**):

$$\chi^2_n = \sum_i [(T_i^{exp} - T_i^{th})^2/(\sigma_i^{exp})^2] + [(NOE^{exp} - NOE^{th})/(\sigma_{noe}^{exp})^2 \quad (11)$$

where $T_1^{th}$, $T_2^{th}$, and $NOE^{th}$ depend on $S^2$, $\tau_e$, and $R_{ex}$ via **Eqs. 1–3** and **5** (*see* **Fig. 4**). The uncertainties of the parameters are estimated using Monte Carlo simulations (typically 200–300). For each residue, five models of increasing complexity are tested, whereas the previously defined diffusion tensor is kept fixed (*see* **Table 2**). For each residue, an appropriate model is selected according to the criteria outlined in **Subheading 3.2., step 10**.
2. After an appropriate model has been found for each residue, all parameters including the diffusion tensor and all parameters of internal dynamics are optimized simultaneously in a final $\chi^2$ minimization. The uncertainties of the free parameters are estimated from a set of Monte Carlo simulations (typically 500).

## 4. Notes

1. For both samples, at this calcium concentration and in the absence of additional salt, both $Ca^{2+}$-binding sites of the pair were deemed saturated, based on the

**Table 2**
**Models of Internal Dynamics for the Analysis of $^{15}N$ Relaxation Data**

| Model | Optimized parameters | Values of fixed parameters |
|---|---|---|
| 1 | $S^2_f$ | $S^2_s = 1$, $\tau_e = 0$, $R_{ex} = 0$ |
| 2 | $S^2_f$, $\tau_e$ | $S^2_s = 1$, $R_{ex} = 0$ |
| 3 | $S^2_f$, $R_{ex}$ | $S^2_s = 1$, $\tau_e = 0$ |
| 4 | $S^2_f$, $\tau_e$, $R_{ex}$ | $S^2_s = 1$ |
| 5 | $S^2_s$, $S^2_f$, $\tau_e$ | $R_{ex} = 0$ |

chemical shifts of aromatic ring protons, which were used to monitor calcium binding previously ***(42)***.

2. At least each series of $T_1$, $T_2$ and the pair of [$^1$H]–$^{15}$N NOE experiments should be acquired in a single session.
3. All pulses should be calibrated carefully and acquisition times in the indirect ($^{15}$N) and direct ($^1$H) dimension should be optimized for adequate resolution and signal-to-noise.
4. A period of 10–20 min of dummy scans preceding each experiment increases the reproducibility of peak intensities.
5. Methods to obtain quantitative information on motions in the μs to ms time-scale include measurement of: $T_2$ as a function of CPMG delay ***(43,44)***, $T_{1\rho}$ as a function $B_1$-field strength ***(14,45)***, $R_{1\rho}-R_1$ as function of $B_1$-field offset ***(46)***, and $T_2$ at various spectrometer field strengths ***(19)***.
6. Because resolution enhancement invariably deteriorates signal-to-noise, well-resolved peaks may be processed with mild resolution enhancement, whereas less well-resolved peaks may be treated with harsher window functions at the expense of signal-to-noise.
7. **Steps 2–7** of **Subheading 3.2.** can be performed with a number of software packages: e.g., with a set of Felix2.3 macros, awk scripts and FORTRAN programs developed by Dr. M. Akke and Dr. A. G. Palmer ***(46)*** (software available from the authors at http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software.html), or using NMRView ***(24)***.
8. There are numerous reasons why a two-parameter fit may be inadequate, such as incorrect estimations of noise contributions using a single spectrum or the comparison of two spectra. This can lead to unreasonably high $\chi^2$ values and subsequent rejection of some fits, especially for high signal-to-noise data.
9. For weak signals, reliable intensity determination can be difficult, especially for longer delay times, which may result in apparent finite offsets. In these circumstances, fitting the data to single exponential decays with finite offsets can lead to statistically significant improvements of the fits.
10. Recently it has been suggested that the magnitude of the CSA may vary substantially between residues ***(47,48)***, but this was not confirmed in a separate study ***(49)***. In the study of cbEGF32-33 the CSA was assumed to be –170 ppm ***(50)***.

11. As can be seen from **Eq. 7**, larger experimental errors yield smaller $\chi^2$ values, leading to the acceptance of simpler models. They are also propagated to produce larger uncertainties of the estimated parameters. This in turn diminishes the significance levels for the acceptance of models of increased complexity. Whereas high signal-to-noise spectra will produce desirably small experimental errors, uncontrollable sources of noise may limit the attainable uncertainties (*see* also **Note 7**).
12. In the absence of structural information, the magnitude of the diffusion tensor may be estimated from the distribution of $T_1/T_2$ ratios ***(51)***.
13. Some authors have suggested the use of entirely dynamics-based selection criteria ***(2)***.

## Acknowledgments

J. M. Werner and I. D. Campbell thank the Wellcome Trust and the OCMS for support. The Oxford Centre for Molecular Sciences is funded by MRC, BBSRC, and EPSRC. A. K. Downing is a Wellcome Trust Senior Research Fellow, and she also thanks the support of the BBSRC and the MRC.

## References

1. Kay, L. E. (1998) Protein dynamics from NMR. *Nat. Struct. Biol.* **5,** 513–517.
2. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by N–15 relaxation using inverse detected 2-dimensional NMR-spectroscopy — the central helix is flexible. *Biochemistry* **31,** 5269–5278.
3. Baldellon, C., Alattia, J. R., Strub, M. P., Pauls, T., Berchtold, M. W., Cave, A., and Padilla, A. (1998) N–15 NMR relaxation studies of calcium-loaded parvalbumin show tight dynamics compared to those of other EF-hand proteins. *Biochemistry* **37,** 9964–9975.
4. Paakkonen, K., Annila, A., Sorsa, T., Pollesello, P., Tilgmann, C., Kilpelainen, I., et al. (1998) Solution structure and main chain dynamics of the regulatory domain (residues 1–91) of human cardiac troponin C. *J. Biol. Chem.* **273,** 15,633–15,638.
5. Gagne, S. M., Tsuda, S., Spyracopoulos, L., Kay, L. E., and Sykes, B. D. (1998) Backbone and methyl dynamics of the regulatory domain of troponin C: anisotropic rotational diffusion and contribution of conformational entropy to calcium affinity. *J. Mol. Biol.* **278,** 667–686.
6. Kordel, J., Skelton, N. J., Akke, M., Palmer, A. G., and Chazin, W. J. (1992) Backbone dynamics of calcium-loaded calbindin-D(9k) studied by 2-dimensional proton-detected N–15 NMR-spectroscopy. *Biochemistry* **31,** 4856–4866.
7. Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D., and Downing, A. K. (2000) Backbone dynamics of a cbEGF domain pair in the presence of calcium. *J. Mol. Biol.* **296,** 1065–1078.
8. Dietz, H. C. and Pyeritz, R. E. (1995) Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders. *Hum. Mol. Genet.* **4,** 1799–1809.

9. Reinhardt, D. P., Mechling, D. E., Boswell, B. A., Keene, D. R., Sakai, L. Y., and Bachinger, H. P. (1997) Calcium determines the shape of fibrillin. *J. Biol. Chem.* **272,** 7368–7373.
10. Cardy, C. M. and Handford, P. A. (1998) Metal ion dependency of microfibrils supports a rod-like conformation for fibrillin-1 calcium-binding epidermal growth factor-like domains. *J. Mol. Biol.* **276,** 855–860.
11. Collodberoud, G., Beroud, C., Ades, L., Black, C., Boxer, M., Brocks, D. J. H., et al. (1998) Marfan Database (third edition): new mutations and new routines for the software. *Nucleic Acids Res.* **26,** 229–233.
12. Kettle, S., Yuan, X. M., Grundy, G., Knott, V., Downing, A. K., and Handford, P. A. (1999) Defective calcium binding to fibrillin-1: consequence of an N2144S change for fibrillin-1 structure and function. *J. Mol. Biol.* **285,** 1277–1287.
13. Abragam, A. (1961) *Principles of Nuclear Magnetism*. Oxford University Press, Oxford.
14. Szyperski, T., Luginbuhl, P., Otting, G., Guntert, P., and Wuthrich, K. (1993) Protein dynamics studied by rotating frame N–15 spin relaxation-times. *J. Biomol. NMR* **3,** 151–164.
15. Akke, M., Liu, J., Cavanagh, J., Erickson, H. P., and Palmer, A. G. (1998) Pervasive conformational fluctuations on microsecond time scales in a fibronectin type III domain. *Nat. Struct. Biol.* **5,** 55–59.
16. Zinn-justin, S., Berthault, P., Guenneugues, M., and Desvaux, H. (1997) Off-resonance rf fields in heteronuclear NMR: Application to the study of slow motions. *J. Biomol. NMR* **10,** 363–372.
17. Peng, J. W. and Wagner, G. (1995) Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple fields. *Biochemistry* **34,** 16733–16752.
18. Farrow, N. A., Zhang, O. W., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral density-function mapping using N–15 relaxation data exclusively. *J. Biomol. NMR* **6,** 153–162.
19. Phan, I. Q. H., Boyd, J., and Campbell, I. D. (1996) Dynamic studies of a fibronectin type I module pair at three frequencies: anisotropic modelling and direct determination of conformational exchange. *J. Biomol. NMR* **8,** 369–378.
20. Kroenke, C. D., Loria, J. P., Lee, L. K., Rance, M., and Palmer, A. G. (1998) Longitudinal and transverse H-1-N-15 dipolar N-15 chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J. Am. Chem. Soc.* **120,** 7905–7915.
21. Woessner, D. E. (1962) Nuclear Spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys.* **37,** 647–654.
22. Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic- resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559.
23. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) Deviations from the simple 2-parameter model-free approach to the inter-

pretation of N–15 nuclear magnetic-relaxation of proteins. *J. Am. Chem. Soc.* **112,** 4989–4991.
24. Johnson, B. A. and Blevins, R. A. (1994) NMR View — a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4,** 603–614.
25. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14,** 51–60.
26. Palmer, A. G., Rance, M., and Wright, P. E. (1991) Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance C-12 heteronuclear NMR-spectroscopy. *J. Am. Chem. Soc.* **113,** 4371–4380.
27. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone dynamics of Escherichia-coli ribonuclease HI- correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246,** 144–163.
28. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by N–15 inverse detected heteronuclear NMR-spectroscopy — application to staphylococcal nuclease. *Biochemistry* **28,** 8972–8979.
29. Farrow, N. A., Zhang, O. W., Formankay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of N–15 longitudinal decay and chemical-exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4,** 727–734.
30. Vold, R. R. and Vold, R. L. (1976) Transverse relaxation in heteronuclear coupled spin systems: AX, AX2, AX3, and AXY. *J. Chem. Phys.* **64,** 320–332.
31. Boyd, J., Hommel, U., and Campbell, I. D. (1990) Influence of cross-correlation between dipolar and anisotropic chemical-shift relaxation mechanisms upon longitudinal relaxation rates of N–15 in macromolecules. *Chem. Phys. Lett.* **175,** 477–482.
32. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T1 and T2 values in proteins. *J. Mag. Reson.* **97,** 359–375.
33. Wang, A. C. and Bax, A. (1993) Minimizing the effects of radiofrequency heating in multidimensional NMR experiments. *J. Biomol. NMR* **3,** 715–720.
34. Noggle, J. H. and Shirmer, R. E. (1971) *The Nuclear Overhasuer Effect: Chemical Applications*. Academic, New York.
35. Grzesiek, S. and Bax, A. (1993) The importance of not saturating H2O in protein NMR — application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115,** 12,593–12,594.
36. Kay, L. E. (1995) Field gradient techniques in NMR-spectroscopy. *Curr. Opin. Struct. Biol.* **5,** 674–681.
37. Skelton, N. J., Palmer, A. G., Akke, M., Kordel, J., Rance, M., and Chazin, W. J. (1993) Practical aspects of 2-dimensional proton-detected N-15 spin relaxation measurements. *J. Mag. Reson. Ser. B* **102,** 253–264.
38. Zhu, G. and Bax, A. (1992) Improved linear prediction of damped NMR signals using modified forward backward linear prediction. *J. Mag. Reson.* **100,** 202–207.

39. Press, W. H., Flannery, B. P., Teukosky, S. A., and Vetterling, W. T. (1990) Numerical Recipes. Cambridge University Press, New York.
40. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D., and Handford, P. A. (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* **85,** 597–605.
41. Copie, V., Tomita, Y., Akiyama, S. K., Aota, S., Yamada, K. M., Venable, R. M., et al. (1998) Solution structure and dynamics of linked cell attachment modules of mouse fibronectin containing the RGD and synergy regions: comparison with the human fibronectin crystal structure. *J. Mol. Biol.* **277,** 663–682.
42. Knott, V., Downing, A. K., Cardy, C. M., and Handford, P. (1996) Calcium binding properties of an epidermal growth factor-like domain pair. *J. Mol. Biol.* **255,** 22–27.
43. Luz, L. and Meiboom, S. (1963) Nuclear magneic resonance study of the protolysis of trimethylammonium ion in aqueous solution — order of reaction with respect to solvent. *J. Chem. Phys.* **39,** 366–370.
44. Loria, J. P., Rance, M., and Palmer, A. G. (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121,** 2331–2332.
45. Deverell, C., Morgan, R. E., and Strange, J. H. (1970) Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. *Mol. Phys.* **18,** 553–559.
46. Akke, M. and Palmer, A. G. (1996) Monitoring macromolecular motions on microsecond to millisecond time scales by R1ρ-R1 constant relaxation-time NMR-spectroscopy. *J. Am. Chem. Soc.* **118,** 911–912.
47. Fushman, D., Tjandra, N., and Cowburn, D. (1998) Direct measurement of N–15 chemical shift anisotropy in solution. *J. Am. Chem. Soc.* **120,** 10,947–10,952.
48. Fushman, D., Tjandra, N., and Cowburn, D. (1999) An approach to direct determination of protein dynamics from N-15 NMR relaxation at multiple fields, independent of variable N-15 chemical shift anisotropy and chemical exchange contributions. *J. Am. Chem. Soc.* **121,** 8577–8582.
49. Kroenke, C. D., Rance, M., and Palmer, A. G. (1999) Variability of the N–15 chemical shift anisotropy in Escherichia coli ribonuclease H in solution. *J. Am. Chem. Soc.* **121,** 10,119–10,125.
50. Tjandra, N., Szabo, A., and Bax, A. (1996) Protein backbone dynamics and N-15 chemical-shift anisotropy from quantitative measurement of relaxation interference effects. *J. Am. Chem. Soc.* **118,** 6986–6991.
51. Clore, G. M., Gronenborn, A. M., Szabo, A., and Tjandra, N. (1998) Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information. *J. Am. Chem. Soc.* **120,** 4889–4890.

# 23

# The Use of Dipolar Couplings for the Structure Refinement of a Pair of Calcium-Binding EGF Domains

**Jonathan Boyd, Iain D. Campbell, and A. Kristina Downing**

## 1. Introduction

The calcium-binding (cb) epidermal growth factor-like (EGF) domain is a common variant of the EGF module that contains a consensus sequence associated with ligation of a single calcium ion ***(1–3)***. Many extracellular proteins include cbEGF domains, and several of these have been associated with human disease (reviewed in **ref. *4***). Two examples from this group are human fibrillin-1 and the low-density liproprotein (LDL) receptor. Moreover, mutations within these proteins have been linked to the Marfan syndrome and familial hypercholesterolaemia, respectively. Recently, we have been using solution NMR to probe the structure, calcium-binding properties and dynamics of tandem cbEGF domains from these two proteins.

The structure of the thirty-second and thirty-third cbEGF domains from human fibrillin-1 (cbEGF32–33) revealed that, in the presence of calcium ions, the two modules adopt a stable rigid rod-like conformation ***(5)***. Analysis of the structure suggests that the domain arrangement is stabilized by the calcium binding to the C-terminal domain in addition to interdomain hydrophobic packing interactions. Furthermore, protein sequence analysis has predicted that the relative orientation of these domains might be a structurally conserved feature of a number of functionally distinct proteins. To test this hypothesis, we have probed the specificity of cbEGF domain packing interactions via investigation of the structure of the cbEGF pair from the LDL receptor (LDLR-AB) (Downing, A. K., et al., manuscript in preparation). For the structure determination of this module pair, we have measured and implemented methods that

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

utilize residual dipolar coupling-derived restraints ***(6–8)***, in addition to employing the usual NMR conformational restraints derived from $^{1}H–^{1}H$ homonuclear nuclear Overhauser enhancements (NOE), $^{3}J_{H\alpha–HN}$ dihedral angles and slow NH exchange.

The use of residual dipolar couplings for structure determination has recently been reviewed ***(9)***. For this chapter, we confine our discussion to a consideration of the dipolar couplings arising between the amide $^{15}N$ nucleus and its directly attached proton from the peptide backbone. In a rigid molecule, the magnitude of the dipole–dipole coupling between the magnetic moments of the $^{1}H$ and $^{15}N$ nuclei depends upon the orientation of the static magnetic field $B_0$ relative to the internuclear vector. It is well established that under normal solution conditions, molecules tumble isotropically, sampling all orientations of conformational space equally. Under these conditions, the dipolar coupling is averaged to zero and does not contribute to the resonance frequencies (although it does strongly influence the relaxation parameters $T_1$, $T_2$ and the $^{1}H–^{15}N$ NOE). However, in a major development, Tjandra and Bax ***(10)*** showed that a relatively small net sample orientation could be introduced when employing a solution mixture containing, in addition to the protein, a completely magnetically oriented nematic uniaxial liquid crystalline phase, known as bicelles ***(11,12)***. More recently, it has been shown that phage or purple membrane may also be used to induce tunable alignment ***(13–15)***. A typical experimental protocol is to adjust the concentration of alignment agent to achieve a net sample orientation of about $10^{-3}$ for the protein. The spectra from the protein then still retain the relative simplicity found for an isotropic solution, with the dipolar coupling reduced from the maximum static molecule value of about 21.6 KHz for an isolated $^{15}N–^{1}H$ spin pair to a manageable range of a few tens of Hz.

NMR structure determination methodology usually relies on identifying, from multidimensional $^{1}H–^{1}H$ NOESY spectra, many local interatomic interactions that are converted to different classes of distance restraint (up to about 5 Å) depending on the intensities of the assigned crosspeaks ***(16)***. Because of the time-averaged nature of NMR data and other sources of inaccuracy such as peak overlap and spin diffusion, there is a tendency for long range error propagation because the structure determination uses predominantly short range $^{1}H–^{1}H$ distance restraints. There have been attempts to reduce the effect of spin diffusion by sample deuteration. It has been demonstrated that distance restraints can be derived from backbone amide protons separated by up to about 7 Å, in a deuterated protein ***(17)***.

In contrast to short-range NOEs, residual dipolar coupling restraints are related to a global principal axis system. In a dilute solution of uniaxial magnetically oriented bicelles the protein molecules still tumble rapidly, but with

unequal orientational probabilities, so that the magnetic dipole–dipole coupling between each pair of $^1$H and $^{15}$N nuclei does not average to zero. The dependence of the residual dipolar coupling on orientational order can be quantified by employing the symmetric traceless molecular alignment tensor, **A** ***(10,18)***, where **A** = 0 for an isotropic system. The alignment tensor is internal to and moves with the protein. Hence, for a sample without significant relative domain or internal motions, the transformation from the principal axes of each $^1$H–$^{15}$N dipolar coupling tensor to the principal axes of the alignment tensor does not depend on time or the details of the molecular motion. In contrast, $^{15}$N–$T_1$, and $T_2$ relaxation rates from isotropic solution depend upon the random motions experienced by the internal molecular rotational diffusion, dipolar and chemical shift tensors with respect to the fixed laboratory magnetic field. The time-dependent motional details can be quantified using the spectral density function. For the system of phospholipids employed in this work, the director of the liquid crystalline phase is perpendicular to the axis of the laboratory magnetic field (*see* **Fig. 1**) and the direct dipolar coupling constant for an isolated $^{15}$N–$^1$H spin pair is given by:

$$^1D_{NH}(Hz) = 1/2\,[S_{NH}\gamma_N\gamma_H h(\mu_o/4\pi^2)\langle r_{NH}^{-3}\rangle]\,\{A_{zz}(3\cos^2\theta - 1)/2 + 0.5(A_{xx} - A_{yy})(\sin^2\theta\cos 2\phi)\} \quad (1)$$

where $A_{ii}$ are the magnitudes of the principal axes of the alignment tensor $A$ and $\theta$ and $\varphi$ are the spherical polar coordinates that specify the orientation of the NH internuclear vector in the principal axis system of the alignment tensor ***(10,19)***. Ottiger and Bax ***(20)*** have suggested that whenever the generalized order parameter ***(21)***, $S_{NH}$, for internal motion is greater than about 0.89 ***(20)***, it is possible to define an effective NH bond length $\langle r_{NH}\rangle = 0.104$ nm with $S_{NH} = 1$.

The degree of protein alignment depends on the concentration of additive (bicelles, phage, or purple membrane) as well as upon the shape of the molecule. Whenever the shape of the protein deviates from axial symmetry, the rhombic term ($A_{xx}$–$A_{yy}$) in **Eq. 1** is expected to assume increased significance. However, it has also been proposed that it will become particularly significant whenever the protein interacts electrostatically and asymmetrically with the additive causing the molecular alignment ***(13)***. The use of residual dipolar coupling-derived restraints has proven to be particularly useful for the structure determination of an elongated molecule, such as the two linked cbEGF domains in LDLR-AB, as described here.

## 2. Materials

1. A purified sample of an $^{15}$N enriched protein, in the concentration range approx 0.2–1.0 m*M* (*see* **Notes 1** and **2**).
2. An NMR spectrometer with a magnetic field strength sufficient to provide good resolution and signal-to-noise ratio in a 2-D $^{15}$N–$^1$H correlated spectrum of the sample.

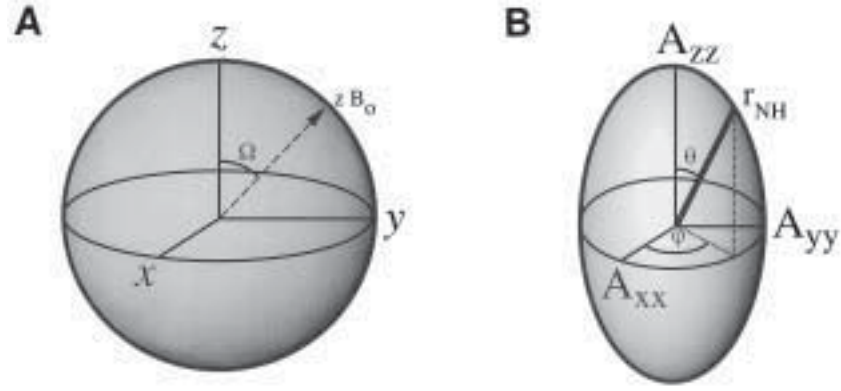


Fig. 1. **(A)** The solute molecules in the oriented phase are subject to a constraint that makes an angle Ω with respect to the direction of the spectrometer magnetic field $B_0$ defined to be the laboratory *z*-axis. In a fully oriented uniaxial nematic liquid crystalline phase, exhibiting cylindrical symmetry about the magnetic field, the anisotropic interactions of the solute are scaled by the factor $(3\cos^2\Omega–1)/2$, where Ω is the angle between the laboratory *z*-axis and the *Z*-axis of the liquid crystalline director reference frame ***(10,37)***. For magnetically oriented diamagnetic bicelles, Ω = 90° and for samples employing purple membranes or bicelles incorporating paramagnetic ions ***(38)***, Ω = 0° and the magnitude of the experimental residual dipolar couplings are scaled, in principle either by –1/2 or 1, respectively. **(B)** The coordinate system employed in calculating the $^1D_{NH}$ residual dipolar couplings. The fixed relative orientation of the principal axis system of the alignment tensor $\mathbf{A}_{ii}$ and the $^{15}N–^1H$ dipolar tensor, shown as the internuclear vector $r_{NH}$, is defined by the polar angles θ and φ.

3. Alignment inducing additives. The bicelle forming phospholipids used here were, 1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC; MW 453.5) and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC; MW 678); they were obtained from Avanti Polar Lipids, Inc. (Alabaster, AL) (*see* **Note 3**).
4. Additives. The protein and bicelle mixtures typically require buffers, salt, and other co-factors; depending upon the pI of the protein, it may be useful to add either CTAB (MW 364.5) or SDS (MW 288.4) to the bicelle solutions to reduce strong electrostatic interactions between the protein and the phospholipids ***(22)***.
5. Computing resources. A high-speed computer with suitable operating system and a Fortran compiler will be required to run the NMR processing software, e.g., Felix 98 (MSI, Inc., San Diego, CA) and X-PLOR ***(23)*** for structure calculations.

## 3. Method

### *3.1. NMR Sample Preparation*

1. DHPC is very hygroscopic and care should be taken to minimize exposure of the sample to the atmosphere; this is helped by solubilizing the DHPC and freeze

drying before weighing. For the samples here, 1 mL of 15% w/v phospholipid solution was made, with $q$ = {DMPC/DHPC} = 2.9, by adding 121.9 mg (179.8 m*M*) of DMPC to 28.1 mg of DHPC (62 m*M*) in an NMR solvent of 5 m*M* Tris-HCl buffer, pH 6.5, 10 m*M* $CaCl_2$, 0.02% $NaN_3$ in 91%/9% $H_2O/^2H_2O$ filtered to 0.45 μm.

2. The 15% w/v phospholipid sample may take approx 24–48 h to solubilize completely; during this period it should be kept at <20°C and agitated periodically.
3. 550 μL of a 5% w/v phospholipid sample, without protein, for $^2H$ NMR tests (*see* **Subheading 3.2., step 2**) is made, at 4°C, by mixing 183 μL of the 15% w/v solution with 367 μL of the NMR solvent (*see* **ref. *1***). The final free concentrations of DMPC and DHPC in the isotropic phase are 59.9 m*M* and 20.7 m*M*, respectively. (Note: It has been suggested ***(24)*** that as much as 5 m*M* free DHPC is likely to remain in the oriented bicelle phase. Vold and Prosser ***(25)*** have argued that the diameter of a bicelle, formed from a solution with $q$ = 3, is about 40 nm, however, more recently, Ottiger and Bax have suggested that the diameter of a typical bicelle will be about 50 nm because of the free DHPC.}
4. A 550 μL 5% w/v protein sample is prepared, as aforementioned, but using 367 μL of a solution containing the $^{15}N$-labeled protein dissolved in the NMR solvent, centrifuged at 176,400 m/s$^2$ for 5 min to remove trace particulate impurities, and then transferred to a clean 5-mm NMR tube.

## 3.2. Testing the Sample Preparation

1. To minimize the possibility of phase separation, it has been recommended ***(20)*** that before the sample is placed in the spectrometer the probe should be preequilibrated to the desired temperature. We have also found it best to keep the protein/phospholipid NMR sample at 4°C when it is not in the spectrometer.
2. The viability of the aligned bicelle phase can be monitored, either with or without the addition of protein, by recording a 1-D $^2H$ spectrum (*see* **Note 4**). In the oriented phase, it is found that the $^2H$ signal from the HOD of the solution appears as a doublet centered at the chemical shift of the $^2H$ nucleus. The splitting of the $^2H$ resonance from a 5% solution ($q$ = 2.9) of ester bicelles was measured to be 9.5 Hz at pH 6.5 and 39°C. The splitting arises from incomplete averaging, in the oriented phase, of the large $^2H$ electric quadrupole interaction of the deuteron in the HOD molecule. For the isotropic solution, this interaction averages to zero and the $^2H$ signal from the HOD molecule appears in the spectrum as a singlet. Although the $^2H$ signal of the oriented phase is composed of two lines, we have encountered no problems with the $^2H$ pulsed free-precession lock system used on our spectrometers, provided that the bicelle system remains stable throughout the course of the experiments.
3. The integrity of the protein in the bicelle solution may be tested by acquisition of a HSQC spectrum ***(26)***. Small changes to the chemical shifts of the $^{15}N$ and $^1H$ resonances occur upon sample orientation, as a result of an additional contribution to the observed resonance frequency from the anisotropic component of the chemical shift tensor of the $^{15}N$ and $^1H$ nuclei ***(27,28)***. Larger shift changes and/

or significant peak broadening may indicate that there is a strong interaction between the protein and the bicelles.

### 3.3. Residual Dipolar Coupling Data Acquisition and Processing

1. NMR pulse sequences. Two RF pulse schemes have been described that have proved to be suitable for measuring $^{15}N$–$^{1}H$ residual dipolar couplings ***(29,30)***. In the first of these methods, the residual dipolar couplings are measured from the $^{1}J_{NH}$ splitting in the $^{15}N$ dimension of a HSQC type experiment incorporating a $S^3E$ pulse sequence element, which is used to select for either the upfield or downfield $^{15}N$ multiplet component in a separate experiment. For the second pulse sequence, known as IPAP, two separate 2-D datasets are recorded, corresponding either to a *J*-coupled in-phase or an antiphase HSQC-type spectrum for the indirect dimension. These datasets are then combined, by addition or subtraction, to form two separate spectra, corresponding to the upfield or the downfield $^{15}N$ multiplet component (*see* **Note 5**).
2. Parameters. Guidelines for data acquisition and processing are suggested based on the experiments recorded on the LDLR-AB module pair. In this study, the first of the pulse sequences aforementioned was used where ($t_1$ $^{15}N$, $t_2$ $^{1}H$). 128 × 2048 complex points were acquired, with acquisition times of 64 and 82 ms, respectively. The data were processed to give a final digital resolution of 0.98 Hz/pt ($F_1$) and 3.05 Hz/pt ($F_2$). Pairs of spectra were collected at 24°C and 39°C.

### 3.4. Data Analysis

1. Residual dipolar couplings $^{1}D_{NH}$ in units of Hz are extracted from the difference in $^{1}J_{HN}$ splitting observed for the datasets recorded in the aligned (39°C) and isotropic (24°C) phases, as illustrated in **Fig. 2** (*see* **Note 6**). In the analysis, we assume that the only dipolar couplings contributing to the experimental $^{1}J_{NH}$ splitting are from the directly attached amide proton (*see* **Note 7**).
2. In order to use residual dipolar couplings as restraints in structure refinement, it is necessary to relate these values to the orientation of the molecular alignment tensor *A*. For this purpose the relationship describing the dipolar coupling $^{1}D_{NH}$, equation 1, may be conveniently recast,

$$^{1}D_{NH}(\theta,\phi) = D_a \{(3\cos^2\theta - 1) + 1.5R(\sin^2\theta \cos 2\phi)\} \quad (2)$$

where $D_a$ in Hz and $D_r$ are the axial and rhombic components of the alignment tensor

$$D_a = 1/3[D_{zz} - (D_{xx} + D_{yy})/2] \text{ and } D_r = 1/3[D_{xx} - D_{yy}] \quad (3)$$

*R*, the rhombicity defined as $D_r/D_a$, is always positive and employing this definition will vary between 0 and 2/3 ***(8)***, and $D_a = (1/2\, S_{NH}\gamma_N\gamma_H{}^{h}[\mu_o/8\pi^2]\langle r_{NH}{}^{-3}\rangle)A_{zz}$ is the axial component of the alignment tensor *A*. It has been recommended that residual dipolar coupling restraints should only be implemented for residues in well-ordered regions of the molecule, where $S^2 > 0.8$. In the ideal case, with many residual dipolar couplings sampling a wide range of orientations with respect to

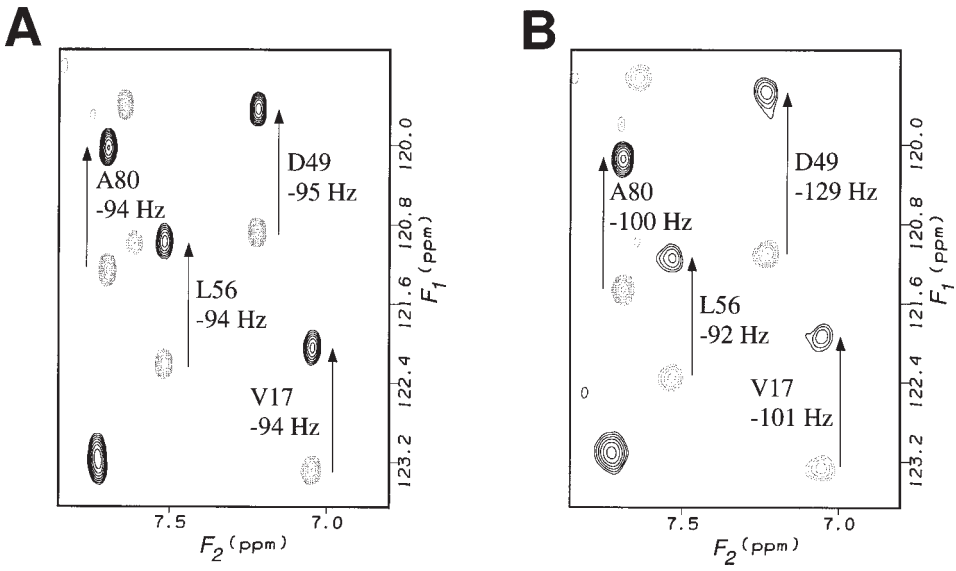


Fig. 2. Selected regions of HSQC-type spectra acquired for the measurement of residual dipolar couplings using spin-state selective excitation (30, *see* text for acquisition parameters). The spectra were recorded at **(A)** 24°C and **(B)** 39°C, in isotropic and aligned phases, respectively. $^1D_{NH}$ splittings for labeled residues are indicated.

the alignment tensor, a histogram of the dipolar couplings will have the same shape as a chemical shift anisotropy (CSA) powder pattern ***(31)***. A histogram of the experimental $^1D_{NH}$ values for the LDLR-AB pair is plotted in **Fig. 3**.

3. Since $D_{zz} + D_{yy} + D_{xx} = \theta$, it follows from **Eqs. 2** and **3** that:

$$D_{zz} = 2D_a;\ D_{yy} = -D_a\,(1 + 1.5R);\ D_{zz} = -D_a\,(1 - 1.5R) \qquad (4)$$

Experimentally, values for $D_{zz}$ and $D_{yy}$ are obtained by taking the average of the high and low extreme values in the histogram, respectively, such that the standard deviations in the estimated values for these parameters are equal to the measurement error. The value for $D_{xx}$ corresponds to the most populated value in the histogram of the observed residual dipolar couplings.

Because the number of dipolar couplings that has been measured for the LDLR-AB pair is limited, estimates for $D_{xx}$, $D_{yy}$, and $D_{zz}$ have been obtained from the histogram and then refined using a grid search procedure ***(8)*** (*see* **Subheading 3.5., step 3**). The cbEGF domain pair structure is highly anistropic and axially symmetric ***(5,32)***, therefore, the probability of sampling vectors that lie parallel to $D_{zz}$ small. Hence, it is anticipated that the maximum observed value for $^1D_{NH}$ underestimates the value of $D_{zz}$. An estimate for $D_{yy}$ corresponds to the minimum value for $^1D_{NH}$ = –28 Hz. The most populated value in the histogram – 17 Hz is taken as $D_{xx}$, hence, $D_{zz}$ is approx 44 Hz.

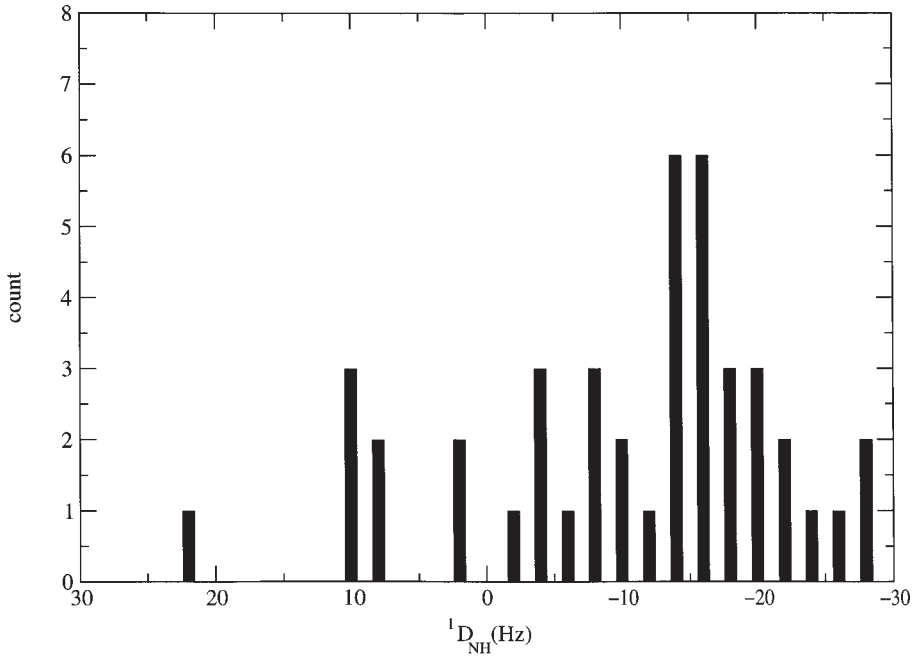


Fig. 3. Histogram of $^1D_{NH}$ values measured for the LDLR-AB cbEGF domain pair.

4. Values for $D_a$ and $R$ are obtained by least squares minimization of **Eq. 4** and using the estimates of $D_{xx}$, $D_{yy}$, and $D_{zz}$ found for LDLR-AB yields values of 23.4 Hz and 0.17, respectively.

### 3.5. Structure Refinement

1. A modified simulated annealing protocol incorporating residual dipolar coupling-derived restraints has been described by Tjandra et al. *(7)* and may be obtained from the authors. In this method, each N–H bond vector is constrained relative to a molecular axis system with axial and rhombic components $D_a$ and $R$. Input data for the procedure comprise estimates for $D_a$ and $R$, as well as a table of restraints in the format shown in **Fig. 4**. Each restraint relates the residual dipolar coupling of residue $n$ to the orientation of the molecular axis system, defined here as residue 500. The molecular axis system is contained in a separate PDB file. It contains four atoms with coordinates corresonding to $x$, $y$, $z$, and the origin. The origin of the molecular axis system is fixed, and its orientation is allowed to float during the simulated annealing protocol, such that it adopts an alignment in best agreement with the experimental measurements. In order to avoid steric clashes, the origin of the axis system should be localized far away (≥50 Å) from the protein.
2. Agreement with experimentally derived restraints is achieved via minimization the term $E_{\text{dipolar}}$ *(7)*:

```
assign ( resid 500  and name OO  )
       ( resid 500  and name Z   )
       ( resid 500  and name X   )
       ( resid 500  and name Y   )
       ( resid   6  and name N   )
       ( resid   6  and name HN  ) -17.2450  0.2000

assign ( resid 500  and name OO  )
       ( resid 500  and name Z   )
       ( resid 500  and name X   )
       ( resid 500  and name Y   )
       ( resid   7  and name N   )
       ( resid   7  and name HN  ) -11.1920  0.2000

assign ( resid 500  and name OO  )
       ( resid 500  and name Z   )
       ( resid 500  and name X   )
       ( resid 500  and name Y   )
       ( resid   8  and name N   )
       ( resid   8  and name HN  )  10.3140  0.2000

assign ( resid 500  and name OO  )
       ( resid 500  and name Z   )
       ( resid 500  and name X   )
       ( resid 500  and name Y   )
       ( resid   9  and name N   )
       ( resid   9  and name HN  ) -19.3740  0.2000
```

Fig. 4. Example of the input format for residual dipolar coupling-derived restraints ***(7)*** in X-PLOR ***(23)***. Residue 500 OO, X, Y, Z correspond to the origin, *x*-, *y*-, and *z*-axes of the alignment tensor axis system, respectively. Restraints are made with reference to the N–HN bond vector for each residue. The value in the second column from the right is the measured residual dipolar coupling. The far right number (0.2000) corresponds to the experimental measurement error, however this is not taken into consideration in the simulated annealing protocol.

$$E_{\text{dipolar}} = k_{\text{dipolar}} ({}^1D_{NH,\text{calc}} - {}^1D_{NH,\text{obs}})^2 \quad (5)$$

where $k_{\text{dipolar}}$ is a force constant, and ${}^1D_{NH,\text{calc}}$ and ${}^1D_{NH,\text{obs}}$ are calculated and observed values of the residual dipolar couplings. This force constant is scaled in tandem with the NOE force constant during the cooling phase of the simulated annealing protocol. The value for $k_{\text{dipolar}}$ is chosen empirically, such that the difference between the calculated vs experimental values for ${}^1D_{NH}$ approximates the measurement error.

3. If the number of measured ${}^1D_{NH}$ values is limited, as in the case of the LDLR-AB pair, estimates for $D_a$ and $R$ may be refined using a grid search method ***(8)***. We have used this method to optimize the two parameters independently. The procedure involves calculating series of structure ensembles (20 structures per ensemble) with varying values for either $D_a$ and $R$. For the 10 lowest total energy structures for each ensemble, estimates for $D_a$ and $R$ were evaluated based on

total and NOE energies. $D_a$ was incremented from 22.5 to 25.0 Hz in steps of 0.5 Hz, and $R$ was incremented from 0.12 to 0.20 in steps of 0.02. For the LDLR-AB pair, optimal agreement was found with $D_a$ 23–23.5 Hz and R 0.16–0.18. These values are consistent with the estimates obtained based on analysis of the histogram shown in **Fig. 3**.

4. Introduction of residual dipolar coupling-derived restraints may result in violation of a subset of the NOE restraints. This is caused by inaccuracies in the NOE-derived distances because of spin diffusion, time averaging, or spectral overlap. In this case, the NOE restraint list may be refined iteratively against the residual dipolar coupling restraints, until the structures satisfy selection criteria. For the LDLR-AB pair, structures shown in **Fig. 5** were chosen with no NOE restraint violated by more than 0.5 Å, no dihedral angle restraint violated by more than 5°, and $|^1D_{NH,\mathrm{calc}} - {}^1D_{NH,\mathrm{exp}}| < 2$ Hz.

### 3.6. Validation of Structure Refinement

1. Although chemical shifts are dependent on molecular alignment, changes in $^{13}$C' chemical shifts ($\Delta\delta^{13}$C'), which are particular sensitive to this effect, have been used to evaluate protein structures before and after refinement against residual dipolar coupling-derived restraints via use of a quality ($Q$) factor defined as:

$$Q = \mathrm{rms}(\Delta\delta^{\mathrm{meas}} - \Delta\delta^{\mathrm{pred}})/\mathrm{rms}(\Delta\delta^{\mathrm{meas}}) \quad (6)$$

where $\Delta\delta$ is the change in chemical shift observed when shifting from an isotropic to an aligned medium ***(28)***. In the absence of $^{13}$C chemical shift data for the LDLR-AB pair, we have chosen instead to evaluate the improvement in structures based on $T_1/T_2$ ratios, which were not incorporated into the structure refinement. The $T_1/T_2$ ratio depends on the angle that NH bond vector angle with respect to the the diffusion tensor (*see* Chapter 22 in this volume). Therefore a comparison of $(T_1/T_2)^{\mathrm{calc}} - (T_1/T_2)^{\mathrm{meas}}$, via analysis of $\chi^2$ values for the structural ensembles computed before and after refinement against the residual dipolar coupling-derived restraints, should directly probe changes in the accuracy of the structure determination. $\chi^2$ is defined as:

$$\chi^2 = \sum_i \{[(T_1/T_2)^{\mathrm{meas}} - (T_1/T_2)^{\mathrm{calc}}]_i / \mathrm{err}\,[(T_1/T_2)^{\mathrm{meas}}]_i\}^2 \quad (7)$$

where $i$ is summed over values for residues that do not manifest large amplitude internal motions or conformational exchange line broadening. For the LDLR-AB pair, this subset of residues was defined as those localized to β-strands excluding residues affected by either fast or slow motions ($\{^1$H$\}-^{15}$N-NOE > 0.7 and $T_2$ > 90 ms). As expected, the LDLR-AB families of structures were best-fit using a prolate ellipsoid model with $D_{//} = D_{zz}$ and $D_\perp = D_{xx} = D_{yy}$ and $D_{//} > D_\perp$, where $D$ is defined as the molecular rotational diffusion tensor. Twenty structures were selected from ensembles calculated before and after residual dipolar coupling based refinement based on total energies and agreement with experimental constraints. The average $\chi^2$ for each family dropped from and average value of 120 to 109 upon refinement, an approx 9% improvement.


Fig. 5. Ensemble of 20 final structures for LDLR-AB, after refinement against residual dipolar coupling-derived restraints. Structures are shown superposed based on the backbone (N,$C^{\alpha}$,C') coordinates of residues 33–67 of the lowest total energy structure, which is highlighted by a backbone ribbbon. This figure was produced using Insight v98.0 (MSI, Inc., San Diego, CA).

2. When studying multidomain constructs, the validity of refinement using a single axis system vs an alignment tensor for each domain must also be justified, as molecular alignment is expected to be affected by the rigidity of interdomain linkages. The linkage between the two domains of the LDLR-AB pair was assessed as rigid based on the small variation in $T_1$, $T_2$, and {$^1$H–$^{15}$N}-NOE values throughout the central region of the construct, together with a correlation time and anisotropy for the domain pair of 4.9 ns and 1.44, respectively, for the structures calculated without the use of residual dipolar coupling derived constraints. These values are in good agreement with those obtained for a homologous pair of cbEGF domains from human fibrillin-1, 5.3 ns and 1.55, which was found to have a rigid interdomain linkage ***(32)***.

## 4. Notes

1. In many cases, to provide sample stability, it has been found necessary to add quite significant quantities of salts, leading to solutions with a high sample conductivity.
2. Guidelines are given for the preparation of samples with a total volume of 550 μL although the use of Shigemi NMR tubes (Shigemi Co., Ltd, Tokyo, Japan) will enable smaller volumes containing less total protein to be used.
3. Although these particular DMPC:DHPC bicelle preparations are only stable above pH 6.0, other phospholipids, containing ether rather than ester linkages ***(33)***, have successfully been used at lower pH.
4. If a bicelle solution does not orient initially at high temperature, repeated cooling and warming of the sample may improve its behavior.
5. If the datasets are recorded from a spectrometer that uses the highest currently available magnetic field strength (> = 17.6 T) then the intensity of the upfield low frequency component of a $^{15}$N $^1J_{NH}$ doublet can be significantly reduced compared to the downfield high frequency component. This is caused by cross-correlated relaxation between the $^{15}$N–$^1$H dipolar and $^{15}$N chemical shift anisotropy relaxation mechanisms. The differential line broadening is expected to increase as the magnetic field strength, the molecular weight, or rotational diffusion anisotropy increase. In these situations it is still possible to record two distinct $^{15}$N–$^1$H spectra to measure the residual dipolar coupling. In the first experiment, the downfield low frequency narrow component can be selected using TROSY ***(34,35)***, and for the second, a HSQC dataset is recorded ***(26)***. The residual dipolar coupling can be measured from the difference between the $^{15}$N chemical shifts in the HSQC and TROSY datasets, recorded for the isotropic and oriented phases. In this case, the contributions to the $^{15}$N resonance frequency from the chemical shift anisotropy and the dynamic frequency shifts should cancel.
6. In some cases, when the molecular alignment is fairly large, the peaks in the spectrum from the oriented phase may be weak or missing. Ignoring relaxation effects this may occur whenever the dipolar couplings become significant compared to the value for the $^1J_{NH}$ coupling in a protein of about –92 Hz. For a HSQC

experiment, recorded from the isotropic phase, the optimum delay for the INEPT preparation stage is about $0.25/{}^1J_{NH}$ s. When dipolar couplings are included, the optimum time will be about $0.25/\{{}^1J_{NH} + |{}^1D_{NH}|\}$ s, and it is not possible to satisfy this relationship uniformly when there is a wide range of $|{}^1D_{NH}|$ values. In this situation, it has been suggested that the IPAP method is less sensitive to the variation in the ${}^1J_{NH}$ splitting ***(29)***. Alternatively, it may be desirable to tune the concentration of media (e.g., bicelles) to optimize the size of the peak splittings. In this study, final measurements were made using a concentration of 3% w/v bicelles. Bicelle solutions at concentration <3% w/v are relatively unstable.

7. When using a fully protonated protein, without any $^2$H isotopic enrichment, other possible $^{15}$NH dipolar interactions, such as from ${}^2D_{NH\alpha}$ and ${}^3D_{NH\beta}$, or from other side-chain protons in close proximity, have not been considered. The ${}^2J_{NH\alpha}$ and ${}^3J_{NH\beta}$ scalar couplings are usually much less than the observed $^{15}$N linewidths. However, the distance between a $^{15}$N and the H$\alpha$ proton is typically around 2.1 Å, which means that this dipolar interaction may only be reduced by about an order of magnitude, compared to that from the directly attached proton. In the situation where the maximum dipolar coupling for the directly attached proton is around 30 Hz, this additional dipolar contribution may be larger than the $^{15}$N linewidth for some of the $^{15}$N resonances with the largest dipolar couplings and may become apparent as an additional splitting. This is unlike the situation that occurs in $^{15}$N-relaxation analyses where the effect of remote protons is typically neglected because of the $r^{-6}$ dependence upon distance, although these protons may still contribute about 2% to the total dipolar relaxation processes. We further assume that the anisotropy of the ${}^1J_{NH}$ scalar coupling is quantitatively negligible ***(36)***.

## Acknowledgments

J. Boyd and I. D. Campbell thank the Oxford Centre for Molecular Sciences, funded by MRC, BBSRC and EPSRC, for support. A. K. Downing is a Wellcome Trust Senior Research Fellow and thanks the support of the BBSRC and the MRC. The authors are grateful to M. Clore for X-PLOR scripts used in structure refinement.

## References

1. Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G. and Campbell, I. D. (1990) The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO J.* **9,** 475–480.
2. Mayhew, M., Handford, P., Baron, M., Tse, A. G. D., Campbell, I. D., and Brownlee, G. G. (1992) Ligand requirements for $Ca^{2+}$ binding to EGF-like domains. *Protein Eng.* **5,** 489–494.
3. Rees, D. J. G., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J., and Brownlee, G. G. (1988) The role of beta-hydroxyaspartate and adjacent carboxylate residues in the 1st EGF domain of human Factor-IX. *EMBO J.* **7,** 2053–2061.

4. Downing, A. K., Handford, P. A., and Campbell, I. D. (2000) Calcium binding EGF-like domains. *Topics Biol. Inorg. Chem.* **3,** 83–99.
5. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D., and Handford, P. A. (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* **85,** 597–605.
6. Bax, A. and Tjandra, N. (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *J. Biomol. NMR* **10,** 289–292.
7. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., and Bax, A. (1997) Use of dipolar H-1-N-15 and H-1-C-13 couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Str. Biol.* **4,** 732–738.
8. Clore, G. M., Gronenborn, A. M., and Tjandra, N. (1998) Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131,** 159–162.
9. Prestegard, J. H. (1998) New techniques in structural NMR — anisotropic interactions. *Nature Str. Biol.* **SS5,** 517–522.
10. Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278,** 1111–1114.
11. Sanders, C. R. and Landis, G. C. (1995) Reconstitution of membrane-proteins into lipid-rich bilayered mixed micelles for NMR-studies. *Biochemistry* **34,** 4030–4040.
12. Sanders, C. R. and Schwonek, J. P. (1992) Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry* **31,** 8898–8905.
13. Sass, J., Cordier, F., Hoffmann, A., Cousin, A., Omichinski, J. G., Lowen, H., and Grzesiek, S. (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. *J. Am. Chem. Soc.* **121,** 2047–2055.
14. Koenig, B. W., Hu, J. S., Ottiger, M., Bose, S., Hendler, R. W., and Bax, A. (1999) NMR measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments. *J. Am. Chem. Soc.* **121,** 1385–1386.
15. Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Str. Biol.* **5,** 1065–1074.
16. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
17. Mal, T. K., Matthews, S. J., Kovacs, H., Campbell, I. D., and Boyd, J. (1998) Some NMR experiments and a structure determination employing a {N-15,H–2} enriched protein. *J. Biomol. NMR* **12,** 259–276.
18. Saupe, A. Z. (1968) Recent results in the field of liquid crystals. *Angew. Chem. Internat. Edit.* **7,** 97–112.
19. Buckingham, A. D. and McLauchlan, K. A. (1967) High resolution nuclear magnetic resonance in partially oriented molecules. *Progr. Nucl. Magn. Reson. Spectrosc.* **2,** 63–110.

20. Ottiger, M. and Bax, A. (1998) Determination of relative N-H-N N-C, C-alpha-C, andC(alpha)-H-alpha effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120,** 12,334–12,341.
21. Lipari, G. and Szabo, A. (1989) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559.
22. Ramirez, B. E. and Bax, A. (1998) Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. *J. Am. Chem. Soc.* **120,** 9106–9107.
23. Brünger, A. T. (1996) *X-PLOR v3. 851*, Yale University, New Haven, CT.
24. Ottiger, M. and Bax, A. (1998) Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *J. Biomol. NMR* **12,** 361–372.
25. Vold, R. R. and Prosser, R. S. (1996) Magnetically oriented phospholipid bilayered micelles for structural studies of polypeptides. Does the ideal bicelle exist? *J. Magn. Reson. Ser. B* **113,** 267–271.
26. Bodenhausen, G. and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69,** 185–189.
27. Boyd, J. and Redfield, C. R. (1999) Characterisation of $^{15}$N chemical shift anisotropy from orientation-dependent changes to $^{15}$N chemical shifts in dilute bicelle solutions. *J. Am. Chem. Soc.* **121,** 7441–7442.
28. Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120,** 6836–6837.
29. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131,** 373–378.
30. Meissner, A., Duus, J. O., and Sorensen, O. W. (1997) Spin-state-selective excitation. Application for E. COSY-type measurement of J(HH) coupling constants. *J. Magn. Reson.* **128,** 92–97.
31. Clore, G. M., Gronenborn, A. M., and Bax, A. (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133,** 216–221.
32. Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D., and Downing, A. K. (2000) Backbone dynamics of a cbEGF domain pair in the presence of calcium. *J. Mol. Biol.* **296,** 1065–1078.
33. Ottiger, M. and Bax, A. (1999) Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *J. Biomol. NMR* **13,** 187–191.
34. Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94,** 12,366–12,371.

35. Schulte-Herbruggen, T., Briand, J., Meissner, A., and Sorensen, O. W. (1999) Spin-state-selective TPPI: a new method for suppression of heteronuclear coupling constants in multidimensional NMR experiments. *J. Magn. Reson.* **139,** 443–446.
36. Kowalewski, J. (1977) Calculations of nuclear spin-spin coupling constants. *Prog. NMR Spectrosc.* **11,** 1–78.
37. Emsley, J. W. (1996) Liquid crystals: general considerations, in *Encyclopedia of Nuclear Magnetic Resonance* (Grant, D. M. and Harris, R. K., eds.), Wiley, London, pp. 2788–2799.
38. Prosser, R. S., Volkov, V. B., and Shiyanovskaya, I. V. (1998) Novel chelate-induced magnetic alignment of biological membranes. *Biophys. J.* **75,** 2163–2169.

# 24

## Vector Geometry Mapping

*A Method to Characterize the Conformation of Helix-Loop-Helix Calcium-Binding Proteins*

**Kyoko L. Yap, James B. Ames, Mark B. Swindells, and Mitsuhiko Ikura**

### 1. Introduction

Members of the EF-hand protein superfamily ***(1)*** share a common calcium-binding *helix-loop-helix* motif as a building block, whose conformation essentially determines biological function. It has been well demonstrated that specific binding of $Ca^{2+}$ to the loop alters conformation of the motif, involving rearrangement of the two helices of the EF-hand in three-dimensional (3-D) space (reviewed in **refs. *2–4***). In $Ca^{2+}$-sensor proteins within this superfamily, the $Ca^{2+}$-induced conformational change is responsible for the sensor activity ***(2)***. For many years this change has been quantitatively characterized by the interhelical angle measured between the two helices ***(5–9)***. Recently, Nelson and Chazin ***(10)*** reported an interaction-based analysis for examining conformational change in EF-hand proteins, including computation of distance difference matrices (calculated between each pair of $C_\alpha$ atoms in two structures). Both methods have advantages and disadvantages. The former approach gives a single, descriptive parameter for a given EF-hand, but is obviously insufficient to describe the conformation and its change in detail. The latter approach is more comprehensive and is sensitive to small conformational changes, but yields a large number of parameters to be interpreted by the user. In this chapter, we describe a method termed Vector Geometry Mapping (VGM), an extension of the "interhelical angle" approach, which produces amore complete and descriptive picture of EF-hand conformations. Providing three angles associated with the two helix vectors of the EF-hand, as well as a simplified 3-D

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

image representation of the vectors, the VGM method permits a more in-depth analysis of the structural diversity observed in the EF-hand protein superfamily ***(11)***. In addition, the method is applicable to proteins containing any multiple-helix structural motif.

## 2. Materials

All calculations are performed by the C program vgm, which is described in **Subheading 3**. Both downloadable and interactive, web-based versions of vgm are available at the web site http://nmr.uhnres.utoronto.ca/ikura/datasoft.html. Requirements for computation and visualization for the downloadable version:

1. For calculation of angles and PDB file generation, a computer capable of running C programs.
2. For visualization, a graphics program that accepts PDB files as input, e.g., Molscript ***(12)*** (available from the web site http://www.avatar.se/molscript/obtain_info.html).
3. A structure containing the EF-hand of interest, in PDB format. The residues that form the EF-hand motif must be known, and very often can be determined by sequence alignment.
4. A copy of vgm: executables for SunOS4/Solaris, HP, Linux, and SGI are available for download; source code for compilation on other platforms can be obtained upon request.

## 3. Methods

### *3.1. vgm Calculation*

vgm superimposes the EF-hands of interest (query EF-hands) on a reference EF-hand using the entering (sequentially first) helix of the EF-hand as the basis for superposition. Angles and distances are calculated based on the position of the exiting (sequentially second) helix of the EF-hand with respect to the position of the entering helix (*see* **Fig. 1**). The program also generates a single PDB file, in which query EF-hands are extracted, superimposed, and positioned in a common coordinate system described in **Subheading 3.1.1.** This PDB-format file can then be used in molecular visualization programs to facilitate simultaneous comparison of conformations of several EF-hands.

The steps described below are executed by vgm and, hence, are transparent to the user.

#### *3.1.1. Determination of the Cartesian Coordinate System*

The common coordinate system in which all EF-hands are superimposed is defined by the reference EF-hand, which is specified by the user. The reference and query EF-hands are represented by straight-line vectors, the end points of which are determined by averaging the structural coordinates of the first or last eleven backbone N, Cα, and C' atoms at either end of the helix (*see* **Note 1**).

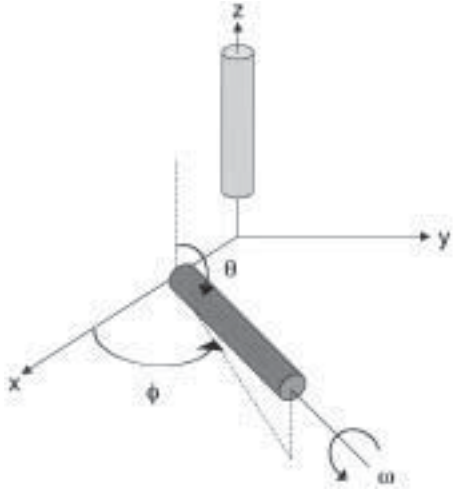


Fig. 1. Vector geometry mapping (VGM) representation of the reference EF-hand. The entering helix vector lies along the *z*-axis and the exiting helix vector "starts" from the *x*-axis.

The entering helix vector of the reference EF-hand defines the position of the +*z*-axis, the position of the *N*-terminal end of the reference exiting helix defines the position of the +*x*-axis, and the intersection of the two axes defines the origin (*see* **Fig. 1**).

### 3.1.2. Superposition of the Query EF-Hands

Each query EF-hand is translated and rotated such that the entering helix vector is aligned with the +*z*-axis, its *C*-terminal end and that of the reference entering helix equidistant from the origin. The EF-hand is rotated about the *z*-axis until itsthe root mean square deviation (RMSD) of the entering helix from the reference entering helix (i.e., the deviation or distance between positions of the backbone N, Cα, and C' atoms) is minimized (*see* **Note 2**). Typically, these RMSD values are well below 1 Å, permitting detailed comparison of the exiting helices ***(11)***. All coordinates from the original PDB files are subject to the same rotation matrix, the new coordinates saved in PDB format.

### 3.1.3. Calculation of Angles and Distances

The geometric position of the exiting helix vector with respect to the entering helix vector is described by three angles. θ is measured between the entering and exiting helix vectors and is 180° less the interhelical angle previously defined ***(13)***. ϕ is measured from the +*x*-axis to the *xy*-projection of the exiting helix vector, counterclockwise about the +*z*-axis. To measure ω, the counterclockwise angle of rotation about the exiting helix vector axis, is measured by rotating the exiting helix vector is translated such that its the exiting helix vector alone it is in the *xz* plane, translating the vector such that the *N*-terminal end is at the same position as the *C*-terminal end of the entering helix (and the EF-hand now forms a "V"), the

modified EF-hand is rotated about the +*z*-axis until the exiting helix vector is in the *xz* plane, and the exiting helix vector alone is rotated about the +*y*-axis by θ degrees. The exiting helix vector, which now lies along the +*z*-axis, is rotated (by ω degrees) until its RMSD from the entering helix is minimized. This angle is useful when two conformational states are compared (*see* **Note 3**).

Distances are calculated between the midpoints of the helix vectors, and between the "outer" end points (i.e., *N*-terminal end of the entering helix, *C*-terminal end of the exiting helix) and the "inner" end points (*C*-terminal end of the entering helix, *N*-terminal end of the exiting helix).

### 3.2. vgm Input

1. Input for the program vgm is a text file that must be of the following format:

   ```
   reference_file.pdb b1 e1 b2 e2 A
   file1.pdb b1 e1 b2 e2
   file2.pdb b1 e1 b2 e2
   ```

   where `reference_file.pdb` is that structure containing the reference EF-hand used for defining the coordinate system; `b1`, `e1`, `b2`, `e2` are the beginning and end residues for the sequentially first and second helices of the EF-hand, and A is an optional chain identifier (normally present in PDB files containing one or more molecules).
2. All lines following the first should describe different EF-hands, and several EF-hands in the same file can be evaluated by listing each EF-hand on a separate line.
3. All filenames must contain the full path to that particular file if it does not reside in the directory from which the program is called.
4. EF-hands should be aligned by structure and the lengths of both entering and exiting helices must be common to all other EF-hands in the input file, including the reference EF-hand (*see* **Note 4**).

### 3.3. vgm Execution

#### 3.3.1. Angle and Distance Output

1. To calculate the angles and distances described in **Subheading 3.1.3.**, the program can be called with

   ```
   vgm input_file
   ```

   where `input_file` is as described in **Subheading 3.2.** Calculated values are output to screen. Values calculated for $Ca^{2+}$-free and -bound calmodulin ***(8,14)*** are shown in **Table 1**.

#### 3.3.2. PDB Formatted Output

1. The program can be called with

   ```
   vgm input_file coord_file
   ```

   where `coord_file` is the output file that will be created to contain the structural coordinates of each superimposed EF-hand.

2. Any graphics program capable of reading PDB files can use this output to display the superimposed EF-hands. We prefer Molscript (v2.0) for its OpenGL feature (allowing interactive rotation of the coordinate system) and its ability to depict helices as cylinders (*see* **Note 5**).

### *3.3.3. Molscript Input File Generation*

1. To generate an input file for Molscript the program can be called with

   ```
   vgm input_file coord_file mol_file
   ```

   where `mol_file` is the created input file required for Molscript (v2.0). This file can be edited to modify color and style settings. By default, the entering helix is shown in white and the exiting helices are shown in green.
2. Default orientation of the coordinate system is that looking down the +*z*-axis, with the +*y*-axis pointing up and the +*x*-axis pointing to the right. In a study of 88 EF-hands in 30 proteins ***(11)***, a rotation matrix approximate to the following was used:

   ```
   -0.66 0.75 0
   -0.13 -0.12 0.98
   0.74 0.65 0.10
   ```

   This rotation will yield the view illustrated in **Fig. 2**.

## 4. Notes

1. To determine the helix vector end points, the user can choose to average the coordinates of either 10 or 11 atoms. The former may be useful for comparison to previously generated interhelical angles because several studies ***(8,13,15–17)*** have reported this angle using ten atom-averaging. There are 3.6 residues, and hence 10.8 backbone (N, $C_\alpha$, C') atoms per turn (360° around a helical wheel) of an $\alpha$ helix. The eleventh atom (e.g., the $C_\alpha$ atom of the fourth residue from the *N*-terminal end) lies about 333° from the first atom (0°) on the helical wheel. Assuming the bonds between the backbone atoms are approximately the same length, a residue occurs every 360°/3.6 = 100° and a backbone atom every 33° around the wheel. The tenth atom lies 300° from the first atom, while the twelfth atom is in nearly the same position (366° or 6°) as the first atom. Thus, the twelfth atom lies almost directly below the first atom (looking down the N-terminal end of the helix). Choosing to average one less atom (i.e., eleven) should give the closest to an even weighting for the average coordinate and thus the most accurate estimation of a center-point of the helix. For this reason, an averaging over eleven atoms is the default option. (The ten atom option is selected by using the vgm10 binary instead of vgm.) It should be noted that all angles differ by less than two degrees, and distances differ by less than 0.3 Å when averaging over ten atoms, when compared to eleven atom-averaging.
2. The N-terminal end of the reference's exiting helix vector will be on the +x-axis by definition; the exiting helix vectors of the query EF-hands usually are not — their position in the coordinate system is determined solely by their superposition on the entering helix vector of the reference.

**Table 1**
**Angle and Distance Output Calculated for Calmodulin (CaM), PDB Codes 1DMO and 1OSA[a]**

| | EF-hand | ϕ | θ | Δω[b] | Midpoint distance | Inner ends distance | Outer ends distance |
|---|---|---|---|---|---|---|---|
| apo-CaM | 1[c] | 123.1 ± 4.5 | 47.6 ± 2.3 | | 9.3 ± 0.1 | 11.2 ± 0.2 | 9.4 ± 0.2 |
| | 2 | 139.5 ± 6.2 | 47.9 ± 4.5 | | 11.6 ± 0.6 | 12.6 ± 0.3 | 12.4 ± 0.6 |
| | 3 | 105.6 ± 4.7 | 44.2 ± 2.5 | | 10.0 ± 0.2 | 10.6 ± 0.2 | 11.1 ± 0.3 |
| | 4[d] | 110.1 ± 10.6 | 52.5 | ± 6.0 | 11.7 ± 0.9 | 11.3 ± 0.3 | 14.0 ± 1.3 |
| $Ca^{2+}$-CaM | 1 | 109.5 | 88.9 | –41 ± 5 | 13.8 | 10.8 | 19.0 |
| | 2 | 104.5 | 91.7 | – 6 ± 7 | 13.7 | 11.5 | 18.5 |
| | 3 | 106.2 | 78.0 | –51 ± 5 | 13.3 | 11.0 | 17.6 |
| | 4[d] | 111.9 | 88.9 | –66 ± 14 | 12.5 | 11.0 | 16.7 |

[a]Standard deviation for apo-CaM (1DMO) values are due to averaging over 30 NMR structures.
[b]Difference between the ω values of apo- and $Ca^{2+}$-CaM.
[c]In this example, apo-CaM EF1 is the reference EF-hand.
[d]The exiting helix of EF4 at the C-terminus of apo- and $Ca^{2+}$-CaM is partially unwound, affecting angle measurement.

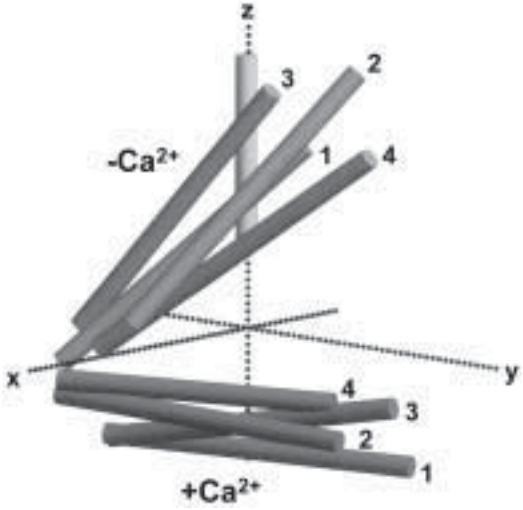


Fig. 2. Example of VGM output for calmodulin (1DMO, 1OSA) as displayed by Molscript ***(11)***. Domains (i.e., an interacting pair of EF-hands) of $Ca^{2+}$-free calmodulin are of the closed conformation; the exiting helices of these EF-hands are labeled "–$Ca^{2+}$." Exiting helices of $Ca^{2+}$-bound, open domain EF-hands are labeled "+$Ca^{2+}$." EF-hands are numbered as they appear in the sequence. Apocalmodulin EF1 is used as the reference in this figure.

3. ω is not a necessarily useful parameter for describing a particular conformation; however, it becomes relevant when the value is compared between two EF-hands that are similarly positioned — either a single EF-hand that undergoes small conformational change (e.g., calpain) or several EF-hands in the same protein (e.g., calmodulin and troponin C). A decrease in ω (negative Δω) between an EF-hand in the $Ca^{2+}$-bound state and in the $Ca^{2+}$-free state indicates that upon binding $Ca^{2+}$, the exiting helix undergoes a clockwise rotation about the helix axis, relative to the position of the entering helix.
4. Alignment by structure rather than sequence alone will yield a more accurate result. Some EF-hands, particularly those situated at the N-terminus of the protein, often have a partially unravelled exiting helix. This is the primary reason for superimposing all EF-hands using the entering helix, which is less prone to structural variation.
5. Cylinder representation in Molscript considers only the structural coordinates of the backbone $C_\alpha$ atoms, compared to the VGM method of averaging the N, $C_\alpha$, and C' atom coordinates to establish vector endpoints. As a result, an entering helix vector calculated by vgm may not lie precisely along the *z*-axis in the Molscript representation.

## Acknowledgements

This work was supported by a grant from the Medical Research Council of Canada (M. Ikura) and the NIH grant EY-12347 (JBA). M. Ikura is a Howard Hughes Medical Institute Research Scholar and a MRCC Scientist.

## References

1. Kawasaki, H., Nakayama, S., and Kretsinger, R. H. (1998) Classification and evolution of EF-hand proteins. *Biometals* **11,** 277–295.

2. Ikura, M. (1995) Calcium binding and conformational response in EF-hand proteins. *Trends Biochem. Sci.* **21,** 14–17.
3. Nelson, M. R. and Chazin, W. J. (1998) Structures of EF-hand $Ca^{2+}$-binding proteins: diversity in the organization, packing and response to $Ca^{2+}$ binding. *Biometals* **11,** 297–318.
4. Gagné, S. M., Li, M. X, McKay, R. T., and Sykes, B. D. (1998) The NMR angle on troponin C. *Biochem. Cell Biol.* **76,** 302–312.
5. Szebenyi, D. M., Obendorf, S. K., and Moffat, K. (1981) Structure of vitamin D-dependent calcium-binding protein from bovine intestine. *Nature* **294,** 327–332.
6. Herzberg, O. and James, M. N. (1985) Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. *Nature* **313,** 653–659.
7. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R., and Cook, W. J. (1985) Three-dimensional structure of calmodulin. *Nature* **315,** 37–40.
8. Zhang, M., Tanaka, T., and Ikura, M. (1995) Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.* **2,** 758–767.
9. Mäler, L. M., Potts, B. C. M., and Chazin, W. J. (1999) High resolution solution structure of apo calcyclin and structural variations in the S100 family of calcium-binding proteins. *J. Biomol. NMR* **13,** 233–247.
10. Nelson, M. R. and Chazin, W. J. (1998) An interaction-based analysis of calcium-induced conformational changes in $Ca^{2+}$ sensor proteins. *Protein Sci.* **7,** 270–282.
11. Yap, K. L., Ames, J. B., Swindells, M. B., and Ikura, M. (1999) Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins* **37,** 499–507.
12. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950.
13. Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C. B., and Bax, A. (1995) Solution structure of calcium-free calmodulin. *Nat. Struct. Biol.* **2,** 768–776.
14. Rao, S. T., Wu, S., Satyshur, K. A., Ling, K. Y., Kung, C., and Sundaralingam, M. (1993) Structure of Paramecium tetraurelia calmodulin at 1.8 Å resolution. *Protein Sci.* **2,** 436–447.
15. Drohat, A. C., Amburgey, J. C., Abildgaard, F., Starich, M. R., Baldisseri, D., and Weber, D. J. (1996) Solution structure of rat apo-S100B(bb) as determined by NMR spectroscopy. *Biochemistry* **35,** 11,577–11,588.
16. Finn, B. E., Evenäs, J., Drakenberg, T., Waltho, J. P., Thulin, E., and Forsén, S. (1995) Calcium-induced structural changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* **2,** 777–783.
17. Gagné, S. M., Tsuda, S., Li, M. X., Smillie, L. B., and Sykes, B. D. (1995) Structures of the troponin C regulatory domain in the apo and calcium-saturated states. *Nat. Struct. Biol.* **2,** 784–789.

# 25

# Use of Calmodulin Antagonists and S-100 Protein Interacting Drugs for Affinity Chromatography

**Ryoji Kobayashi**

## 1. Introduction

A complete understanding of the organizing and functioning of an intracellular $Ca^{2+}$-signaling system requires the cooperation of several different approaches, such as genetic manipulation, biochemistry, cell biology, structural biology, and molecular pharmacology. The advent of specific and effective pharmacological tools is always an event of considerable interest. A large number of $Ca^{2+}$-dependent cellular processes have been revealed over the past 20 yr. However, relationships among each signal pathway are complex and uncertainties concerning the cellular responses in the $Ca^{2+}$-signaling systems remain. Although the biochemistry and molecular biology of $Ca^{2+}$-binding proteins have progressed, it has been more difficult to understand their function in intact cells. For this reason researchers have long sought the development of specific antagonists for each $Ca^{2+}$-binding protein that would permit the definitive determination of the physiological role of the individual $Ca^{2+}$-binding proteins. Calmodulin antagonists, such as phenothiazines and W7, are often used as pharmacological tools to clarify the $Ca^{2+}$/calmodulin-dependent reactions. Calmodulin antagonists of strikingly heterogeneous chemical structure modify the interactions of calmodulin with target proteins. However, a detailed structural description of drug-calmodulin binding is only available in the case of trifluoperazine.

Calmodulin antagonists also bind other $Ca^{2+}$-binding proteins, such as troponin C, brain S-100 proteins, and annexins. $Ca^{2+}$-dependent affinity chromatography using a calmodulin antagonist (phenothiazine) was first demonstrated by Jamieson and Vanaman ***(1)***. Later, Marshak et al. ***(2)*** and Endo et al. ***(3)*** reported

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

that S-100A1 and S-100B also bind to phenothiazine- and W7-Sepharose conjugates in a $Ca^{2+}$-dependent manner and quantitatively compete with calmodulin for binding to the immobilized drugs. They also reported the use of calmodulin antagonist-Sepharose conjugates for the purification of S-100A1 and S-100B. Furthermore, studies concerning the binding of isolated domains of several of these proteins have also been reported by Vogel et. al. ***(4)***. In this chapter, attention is focused on the interaction of newly developed pharmacological agents with nonneuronal S-100 family proteins.

### *1.1. Coupling of Ligands to Matrix for Drug-Affinity Chromatography*

Affinity matrices, coupling solvents, temperature and blocking methods should be chosen based on coupling chemistry ***(5–9)***, solubility and stability of the drug of interest, stability of the matrix in the organic solvent, and the length of spacer arm of the matrix ***(10,11)***.

## 2. Materials

1. Drugs: W7 (*N*-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide) are available from Sigma Co. and Calbiochem Co. Fluphenazine·2HCl (4-[3-[2-(trifluoromethyl)-10H-phenothiazin-10-yl]propyl]-1-piperazine ethanol) was obtained from Research Biochemicals International, Inc. Amlexanox (2-amino-7-(1-methylethyl)-5-oxo-5H-[1]benzopyrano(2,3-b)pyridine-3-carboxylic acid) was a generous gift from Takeda Pharmaceutical Co., Japan. Cromolyn (cromoglycic acid) sodium salt (5,5'-[(2-hydroxy-1,3-propanediyl)bis-(oxy)]bis[4-oxo-4H–1-benzopyran-2-carboxylic acid] disodium salt) can be purchased from Sigma Co. Tranilast (rizaben, *N*-(3',4'-dimethoxycinnamoyl) anthranilic acid) was a generous gift from Kissei Yakuhin Co. (Matsumoto, Japan). Structures of drugs mentioned are in **Fig. 1**.
2. Affinity matrices: AF-amino Toyopearl 650M and Epoxy Toyopearl 650M, hydrophilic vinyl polymer supports were obtained from Tosoh Co. (Tokyo, Japan). These affinity matrices are relatively stable in organic solvents, such as dimethylformamide and dioxane. Epoxy-activated Sepharose 6B was obtained from Amersham Pharmacia Biotech (Uppsala, Sweden).
3. EDC (*N*-ethyl-*N*'-(3-dimethylaminopropyl)carbodiimide hydrochloride).
4. DMF (*N*,*N*'-dimethylformamide).
5. Sodium acetate.
6. Acetic anhydride.
7. Ethanolamine.
8. Dioxane.
9. Rinsing Buffers A and B: 0.1 *M* Tris-HCl buffer, pH 7.6, and 100 mL 20 m*M* Tris-HCl buffer, pH 7.6.
10. Buffer A: 20 m*M* Tris-HCl, 0.1 m*M* EGTA, pH 7.5.
11. Buffer B: 20 m*M* Tris-HCl, 0.5 m*M* $CaCl_2$, pH 7.5.

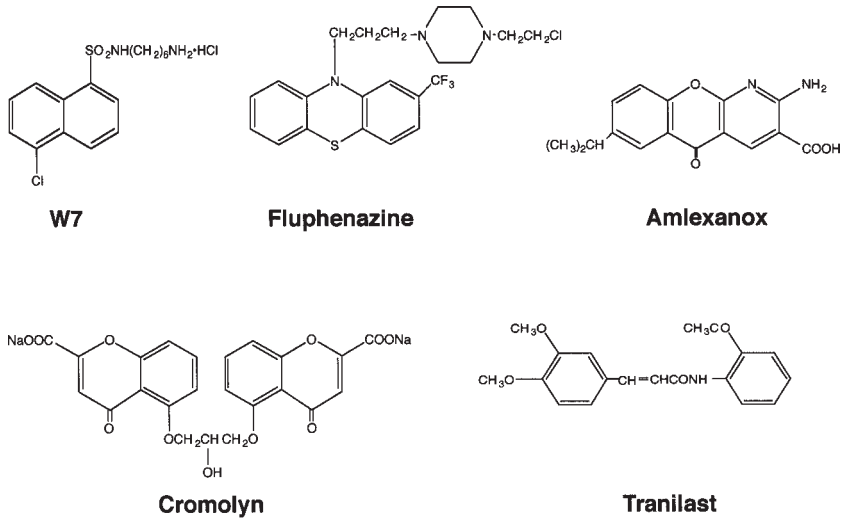


Fig. 1. Structures of substances mentioned.

12. Buffer C: 20 m*M* Tris-HCl, 2 m*M* $CaCl_2$, pH 7.5.
13. Buffer D: 20 m*M* Tris-HCl, 6 *M* urea, pH 7.5.

## 3. Methods

### 3.1. Coupling of Amlexanox to AF-amino Toyopearl 650M

1. Wash the resin (5 g wet weight) with 20 mL of DMF by decantation (three times).
2. Add 0.34 mmol (100 mg) of Amlexanox (dissolved in 1 mL of DMF) to the washed gel.
3. After gentle mixing, add EDC (0.15 g, suspended in 10 mL of DMF) to the suspension. Adjust the pH of the mixture to 5.0 by the dropwise addition of 1 *N* HCl. The pH should be measured using pH paper because organic solvents may damage electrodes.
4. After incubation with gentle shaking for 24 h at room temperature, readjust the pH to 5.0 by the dropwise addition of 1 *N* NaOH or 1 *N* HCl and further incubate for 24 h.
5. Wash the resin successively with 20 mL of DMF (three times) and 50 mL of distilled water (three times) on a sintered glass filter funnel.
6. Block any remaining active groups by incubating the washed resin with 0.2 *M* sodium acetate (4 mL) and acetic anhydride (2 mL) for 30 min at 0°C.
7. Add 2 mL of acetic anhydride to the gel suspension and incubate with gentle shaking further for 30 min at room temperature.
8. Wash the resin successively with distilled water (100 mL), 0.1 *N* NaOH (100 mL) and distilled water (300 mL) with a sintered glass filter funnel. The coupled affinity resin is now ready for use (*see* **Note 1**).

### 3.2. Coupling of Cromolyn to AF-amino Toyopearl 650M

1. Dissolve 0.5 μmol (234 mg) of Cromolyn (free form) in 26 mL of a coupling solution (DMF: $H_2O$ = 25 mL: 1 mL) at 40°C. Adjust pH to 4.0 with the dropwise addition of 1 *N* HCl. The pH should be measured using pH paper.
2. Wash 5 g (wet weight) of AF-amino Toyopearl with 100 mL of the coupling solution on a sintered glass filter funnel.
3. Mix the washed resin and the Cromolyn solution with gentle shaking and then add 0.5 g of EDC in solid.
4. Incubate the mixture with a shaker in a water bath at 40°C for 24 h.
5. After the coupling is completed, wash away excess Cromolyn with 200 mL of the coupling buffer and 100 mL of distilled water on a sintered glass filter funnel.
6. Block any remaining active group as described in **Subheading 3.1.**, **steps 6–8**.

### 3.3. Coupling of Tranilast to AF-amino Toyopearl 650M

1. Dissolve 2.5 μmol (820 mg) of Tranilast in 20 mL of 90% DMF at room temperature. Adjust pH to 4.5 by the dropwise addition of 1 *N* HCl. The pH should be measured using pH paper.
2. Wash 5 g (wet weight) of AF-amino Toyopearl with 100 mL of 90% DMF (pH 4.5) on a sintered glass filter funnel.
3. Mix the washed resin and the Tranilast solution and then add 1 g of EDC in solid.
4. Incubate the mixture for 1 h with gentle shaking and readjust the pH to 4.5 by 1 *N* NaOH or 1 *N* HCl, and further shaken for 24 h at room temperature.
5. Wash the resin successively with 100 mL of 90% DMF (without adjusting pH) and 100 mL of distilled water.
6. Block any remaining active group as described in **Subheading 3.1., steps 6–8**.

### 3.4. Coupling of Fluphenazine to Epoxy-Activated Sepharose 6B

1. Dissolve 200 μmole (102 mg) of Fluphenazine·2HCl in 4.9 mL of a coupling solution (Dioxane: $H_2O$: 1 *N* NaOH = 2 mL: 2 mL: 0.9 mL, pH 10.0).
2. Weight out 0.5 g Epoxy-activated Sepharose 6B (0.5 g freeze-dried powder gives about 1.7 mL of gel) and suspend it in 50 mL of distilled water. Wash swollen gel with 200 mL of distilled water and 20 mL of the coupling buffer on a sintered glass filter funnel.
3. Mix the washed gel and the Fluphenazine solution and incubate with gentle shaking in a water bath at 45°C for 24 h.
4. After the coupling is completed, wash the gel with 100 mL of the coupling solution, pH 10.0, on a sintered glass filter funnel.
5. Any nonreacted groups on the gel should be blocked with gentle shaking in 1 *M* ethanolamine, pH 8.0, for 15 h at room temperature.
6. Wash the product successively with 100 mL of distilled water, and 100 mL of rinsing buffers A and B on a sintered glass filter funnel. The coupled affinity resin is now ready for use.

### 3.5. Coupling of W7 to Epoxy-Activated Sepharose 6B

1. Dissolve 26.5 μmol (10 mg) of W7 in 2 mL of 30% dioxane and adjust pH to 9.8 with the dropwise addition of 1 *N* NaOH. The pH should be measured using pH paper.
2. Weigh out 0.5 g of epoxy-activated Sepharose 6B and suspend it in 50 mL of distilled water. Wash swollen gel with 200 mL of distilled water and 20 mL of 30% dioxane on a sintered glass filter funnel.
3. Mix the washed gel and the W7 solution and incubate with gentle shaking in a water bath at 37°C for 16 h.
4. After the coupling is completed, wash the gel with 20 mL of 30% dioxane, pH 9.8.
5. Nonreacted groups should be blocked by standing in 10 mL of 1 *M* ethanolamine, pH 8.0 for 15 h at room temperature.
6. Wash the coupled resin successively with 100 mL of distilled water and rinsing buffers A and B. The coupled affinity resin is now ready for use.

### 3.6. $Ca^{2+}$-Dependent Affinity Chromatography of S-100 Protein Family on Drug-Coupled Affinity Gels

Affinity chromatography on inhibitor (antagonist)-coupled Sepharose is a widely accepted approach for isolating proteins. Here, we demonstrate the use of antiallergic drug-Toyopearl (or Sepharose) conjugates for the rapid purification of nonneuronal S-100 family proteins.

### 3.7. Example: Identification and Purification of S-100 Family Proteins From Bovine Lung Extract Using Drug-Affinity Chromatography (12,13,14)

#### 3.7.1. Preparation of Bovine Lung Extract for Drug Affinity Chromatography

1. All purification steps were performed at 4°C unless stated otherwise. Homogenize bovine lung (25 g, wet weight) in 6 vol of Buffer A by a Polytron homogenizer (set 9) for 3 min. Centrifuge the homogenate at 15,000*g* for 45 min. Filter the supernatant through glass wool.
2. The filtrate is adjusted to final calcium concentration of 0.5 m*M* by the addition of 1 *M* $CaCl_2$.
3. After stirring for 15 min, centrifuge the solution at 15,000*g* for 45 min, and filter the supernatant through filter paper.

#### 3.7.2. $Ca^{2+}$-Dependent Drug-Affinity Chromatography of Bovine Lung Extract (**Fig. 2**)

1. Preequilibrate a drug-affinity column (1 × 5 cm) with 10–15 vol of Buffer B.
2. Apply the protein extract to the drug-affinity column and wash the column with 100 vol of Buffer B.

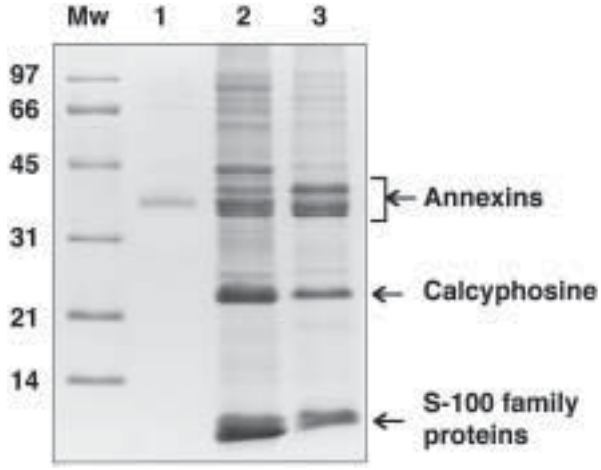


Fig. 2. Tricine/SDS/PAGE ***(15)*** of the proteins obtained from bovine lung extract by Amlexanox- and Cromolyn- affinity chromatographs. The gels were stained with Coomassie brilliant blue R–250. The numbers on the left indicate molecular mass standards (Bio-Rad) in kilo-Daltons. Affinity columns were eluted with an EGTA-containing buffer (Buffer C). Lane 1, acetylated amino-Toyopearl column for control; lane 2, Amlexanox-Toyopearl column; lane 3, Cromolyn-Toyopearl column.

3. Elute the protein with 150 mL of Buffer C (20 m*M* Tris-HCl, 2 m*M* EGTA, pH 7.5) and then with Buffer D. Monitor the column eluate at 280 nm, and collect the fractions corresponding to each major peak.
4. To verify purity of the fraction, subject the column eluate to Tricine/SDS/PAGE ***(15)***, Western-blotting, and reverse phase (RP)-HPLC.

### *3.7.3. Identification and Separation of $Ca^{2+}$-Binding Proteins from a Drug-Affinity Column by RP-HPLC (**Fig. 3**)*

In most instances, the identification and separation of each $Ca^{2+}$-binding proteins (EF-hand proteins) can be archived by analytical RP-HPLC combined with SDS/PAGE. Analytical RP-HPLC should be carried out on a narrow $C_{18}$ column using a 0–60% acetonitrile gradient. Analysis on a Tricine/SDS/PAGE as reported by Schägger and von Jagow ***(15)*** is often helpful in confirming the homogeneity of the preparation.

1. The protein sample obtained from the drug-affinity column could be boiled in a water bath for 3 min to eliminate heat labile proteins. Calmodulin and other EF-hand proteins such as S-100 protein family and calcyphosine are heat stable. Alternatively, protein sample from a drug-affinity column could be separated using an ion-exchange column chromatography, such as Q-Sepharose, DEAE cellulose, and Mono Q-FPLC.

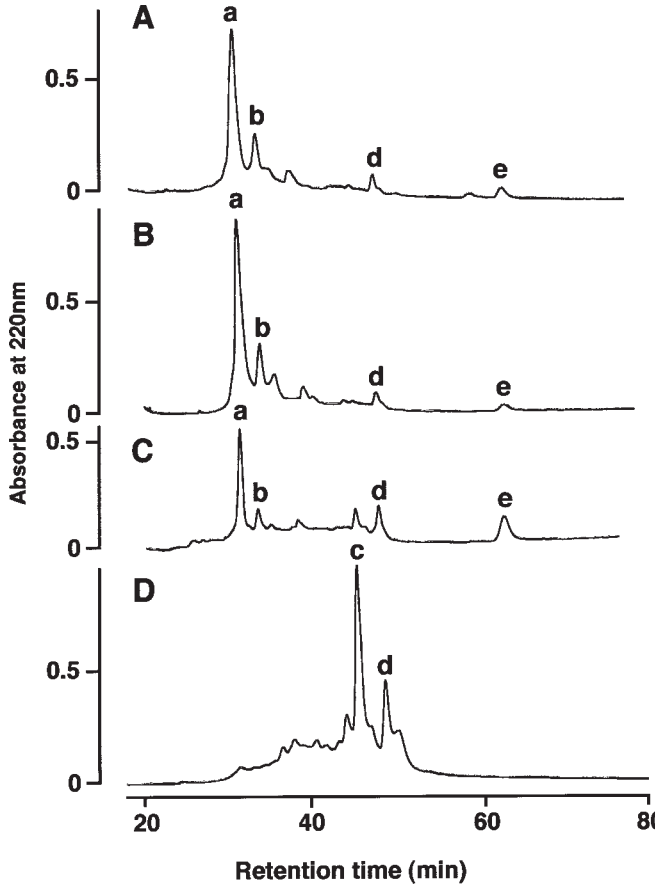


Fig. 3. Elution profiles of $C_{18}$ reverse-phase HPLC of (TSK $C_{18}$ column, Tosoh, Co., Japan) the S-100 family proteins from drug-affinity chromatography of bovine lung extract. **(A)** Tranilast-Toyopearl column; **(B)** Cromolyn-Toyopearl column; **(C)** Amlexanox-Toyopearl column; **(D)** phenyl-Sepharose column. "**a**" and "**b**," S-100S12; "**c**" and "**d**," S-100A2 (S-100L); "**e**," S-100A13. S-100 family proteins (a, b, c, d, and e) were identification by protein sequencing after lysylendopeptidase digested in followed by HPLC separation.

2. Set the UV monitor of HPLC to a wavelength at 220 or 280 nm. Equilibrate a 4.6 × 25 cm TSK $C_{18}$ column (Tosoh, Co., Japan) in aqueous 0.1% TFA at 1 mL/min until a flat baseline is obtained at 220 nm. Clarify the protein sample containing 0.1% TFA by filtration (0.2-μm Teflon filter mounted on a syringe) or centrifugation.
3. Inject suitable amount of the protein preparation that gives a 60–90% full-scale recorder deflection. Initiate a liner gradient from 0–40% acetonitrile containing

0.1% TFA at a flow rate of 1 mL/min (at 13.3%/min for min). Then apply a shallow acetonitrile gradient: 40–64% acetonitrile at 0.34%/min.
4. Collect UV absorbing material using a peak-actuated fraction collector.
5. Lyophilize samples and analyze for purity on Tricine/SDS/PAGE ***(15)***. To identify the protein, subject purified sample to TOF-mass and/or protein sequencing analysis.

### 3.8. Analysis of $Ca^{2+}$-Dependent Interaction of S-100 Family Proteins and their Mutant Proteins with Immobilized Drugs

EF-hand $Ca^{2+}$-binding proteins, such as calmodulin and S-100 protein family interact with calmodulin antagonists in a $Ca^{2+}$-dependent manner. These calmodulin antagonists also inhibit the $Ca^{2+}$-dependent activation of enzymes by calmodulin and therefore are useful probes of the relationships between structure and function in calmodulin and S-100 protein family. Recently, a new class of selective S-100 interacting agents, such as Cromolyn, Amlexanox, and Tranilast (antiallergic drugs), are reported. In this Subheading, we describe the advantage of affinity chromatographic analysis for $Ca^{2+}$-dependent drug-protein interaction (*see* **Note 2**).

#### 3.8.1. Preparation of Standardized Drug-Affinity Columns

1. Pour the slurry of a drug-immobilized matrix (1 mL of bed volume) into Polyprep-Column (Bio-Rad) and equilibrate the column with 10 mL of Buffer B.
2. Apply the protein sample (150 μg) in a small volume of Buffer A and elute the column with 10 mL of the same buffer. Collect each 1 mL of eluate.
3. Elute the bound protein successively with 10 mL of Buffer C and 10 mL of Buffer D. Collect each 1 mL of eluate.
4. To identify the protein, subject each fraction to Tricine/SDS/PAGE ***(15)***.

#### 3.8.2. Example 1: Affinity of Recombinant S100A12 and S100A13 to Amlexanox **(13,14)**

The affinity of recombinant S-100A12 and S-100A13 to the anti-allergic drug, Amlexanox, was examined. The recombinant proteins expressed from the bovine lung cDNA were applied to the Amlexanox-AF amino Toyopearl column (1 mL) with the $Ca^{2+}$ containing buffer (Buffer B). After washing the column with the same buffer, the bound protein was eluted with 2 m*M* EGTA (Buffer C). As shown in **Fig. 4**, the elution pattern and 12% Tricine/SDS/PAGE of recombinant S-100A12 indicated that it bound to Amlexanox in the presence of $Ca^{2+}$, and dissociated from the drug by removing $Ca^{2+}$ from the protein. The recombinant S-100A13 was also examined in a similar manner and it was found that this protein also bound to the drug in a $Ca^{2+}$-dependent manner, although a large part of S-10A13 passed through the column.

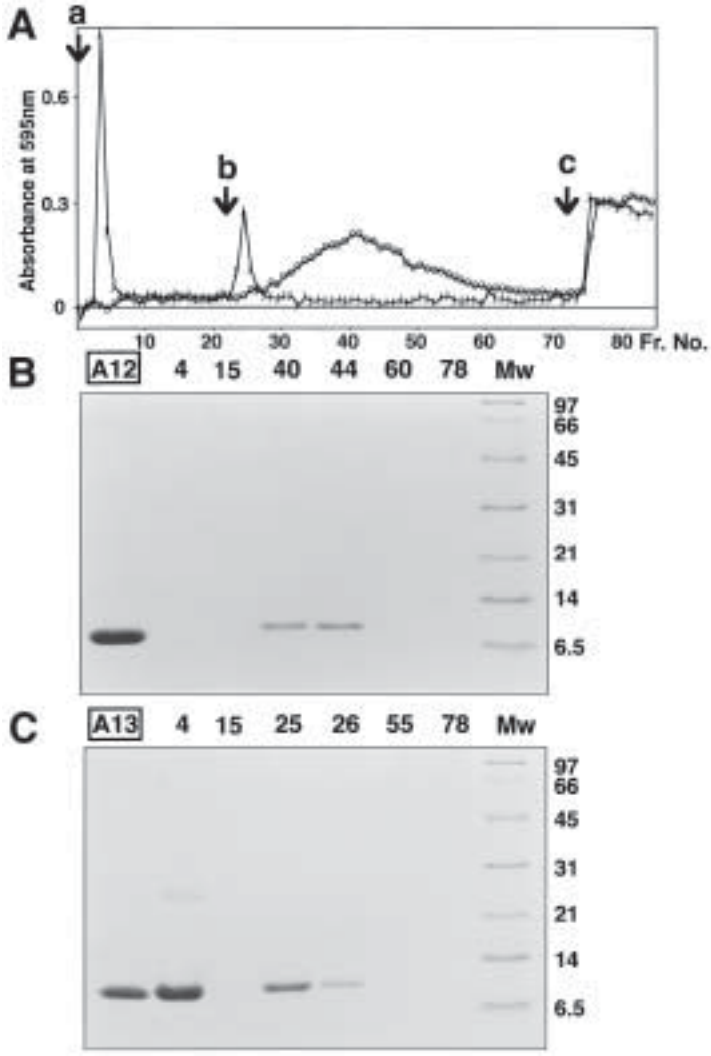


Fig. 4. Affinity chromatography of the recombinant S-100A12 and S-100A13 and Tricine/SDS/PAGE ***(15)*** analysis of selected fractions. **(A)** Elution curves of the recombinant proteins from an Amlexanox-Toyopearl column (o, S-100A12; ı, S-100A13). **a**, Sample loading followed by washing with $Ca^{2+}$-containing buffer (Buffer B); **b**, addition of the buffer containing 2 m*M* EGTA (Buffer C); **c**, addition of the buffer containing 6 *M* urea (Buffer D). **(B)** Tricine/SDS/PAGE (12%) analysis of the recombinant S-100A12 and the fractions (4, 15, 40, 44, 60, and 78) from Amlexanox-Toyopearl column chromatography (*see* **A**). **(C)** 12% Tricine/SDS/PAGE analysis of the recombinant S-100A13 and the fractions (4, 15, 25, 26, 55, and 78) from Amlexanox-Toyopearl column chromatography. Size standards (Mw) are shown in kDa.

Both recombinant S-100A12 and S-100A13 interacted with the drug in a $Ca^{2+}$-dependent manner and the affinity of S-100A12 to Amlexanox was much higher than that of S-100A13. The observation implies that the spatial arrangement of the amino acid residues that interact with the drug differs in S-100A12 and S-100A13.

### 3.8.3. Example 2: Fluphenazine-Sepharose Chromatography of Recombinant and Mutant S-100A1 Proteins

Standardized Fluphenazine-Epoxy activated Sepharose columns of equivalent size and volume (1 mL) as well as fraction size were used to chromatograph equivalent volumes and quantities of recombinant and mutant S-100A1 proteins. Elution profiles and Tricine/SDS/PAGE analyses of the recombinant and mutant S-100A1 proteins are shown in **Fig. 5**. The recombinant S-100A1 protein bound in the presence of $Ca^{2+}$ and was eluted in the presence of EGTA (a $Ca^{2+}$-chelating agent). The C-terminal deletion mutant (ΔFFWENS) of S-100A1 was detected in the $Ca^{2+}$ buffer wash. These results indicate that C-terminal six residues are necessary for $Ca^{2+}$-dependent interaction of S-100A1 with the drug. Lander et al. ***(16)*** and Osterloh et al. ***(17)*** demonstrated that proteins lacking the carboxy-terminal nine residues (especially three hydrophobic residues, Phe-88, Phe-89, and Trp-90) of S-100A1 exhibited reduced $Ca^{2+}$-dependent interaction with the Cap Z peptide, TRTK–12. These result indicate that Fluphenazine interacts with S-100A1 at the target recognition domain in the protein.

### 3.8.4. Example 3: Specificity of Amlexanox-S-100A12 Interaction (**Fig. 6**)

Antiallergic drugs of strikingly heterogeneous chemical structure bind S-100A12 and S-100A13 in a $Ca^{2+}$-dependent manner and can be used for purification of these proteins by a protocol utilizing this $Ca^{2+}$-binding interaction. $Ca^{2+}$-dependent interaction of S-100A1 and S-100B with calmodulin antagonists, such as phenothiazine and W7 was also reported. To analyze specificity of the drug-binding to S-100A12, standardized Fluphenazine-Sepharose and Amlexanox-Toyopearl columns of equivalent size were used. As expected, S-100A12 bound Amlexanox-Toyopearl in the presence of $Ca^{2+}$ (Buffer B) and was eluted in the presence of $Ca^{2+}$-chelating agent, EGTA (BufferC). S-100A12 also bound to Fluphenazine-Sepharose in the presence of $Ca^{2+}$ and eluted in the presence of EGTA. However, a large amount of the protein was detected in the $Ca^{2+}$ buffer wash. The result indicates that S-100A12 has much weaker interaction with Fluphenazine than that with Amlexanox.

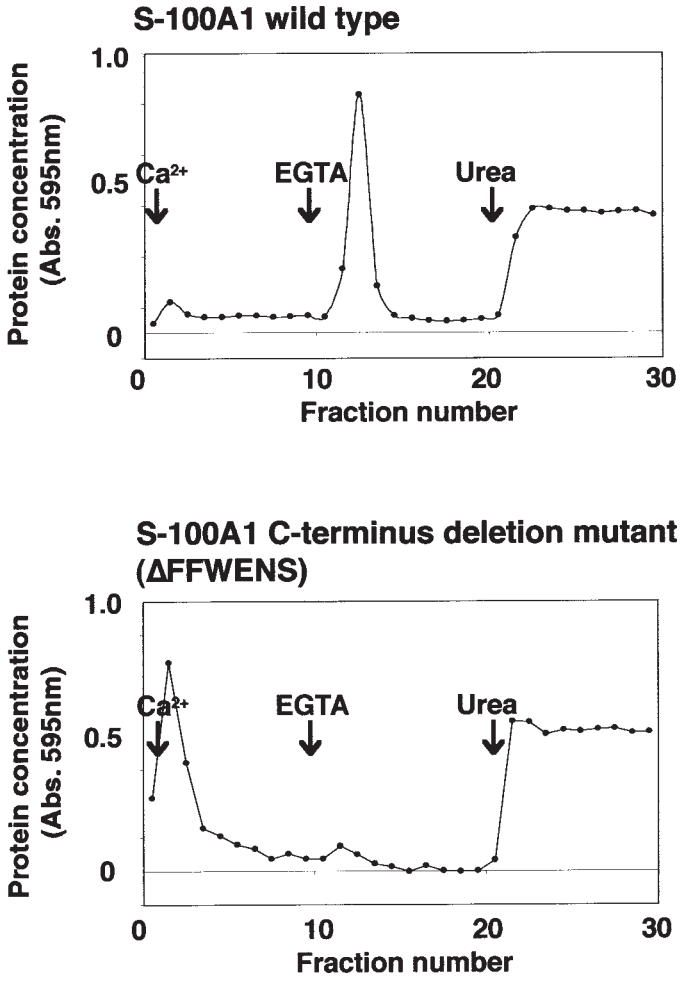


Fig. 5. Affinity chromatography of recombinant S-100A1 (wild-type and *C*-terminal $^{88}$FFWENS$^{93}$ deletion mutant) on a Fluphenazine-Sepharose column. $Ca^{2+}$, sample loading followed by washing with the $Ca^{2+}$-containing buffer (Buffer B); EGTA, addition of the buffer containing 2 m*M* EGTA (Buffer C); urea, addition of the buffer containing 6 *M* urea (Buffer D).

## 4. Notes

1. Epoxy-activated Toyopearl 650M can be used to couple Amlexanox through its amino group. However, the extent of nonspecific binding to the gel is higher than that of Amlexanox coupled AF-amino Toyopearl.

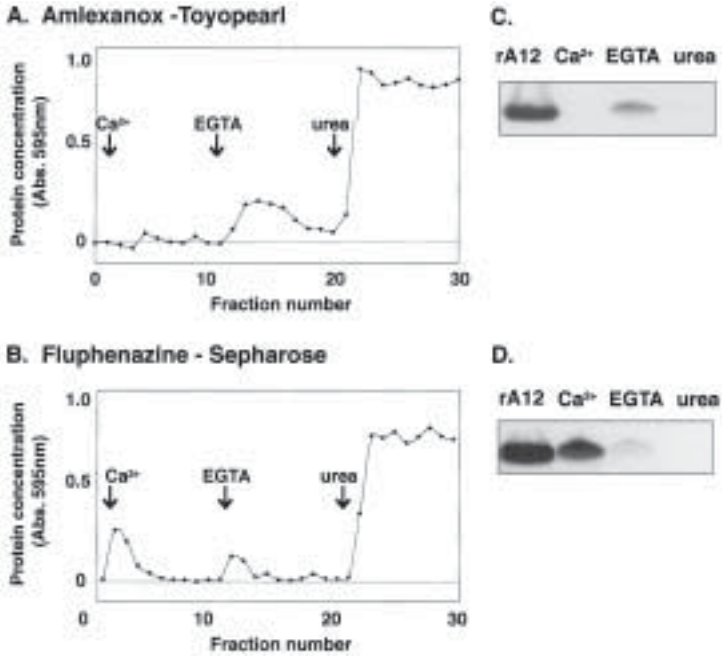


Fig. 6. Affinity chromatography of recombinant S-100A12 on an Amlexanox-Toyopearl column **(A)** or a Fluphenazine-Sepharose column **(B)**. $Ca^{2+}$, sample loading followed by washing with the $Ca^{2+}$-containing (Buffer B); EGTA, addition of the buffer containing 2 m*M* EGTA (Buffer C); urea, addition of the buffer containing 6 *M* urea (Buffer D). C and D, 12% Tricine/SDS/PAGE analyses of the recombinant S-100A12 and the $Ca^{2+}$-wash, EGTA eluate and urea eluate ($Ca^{2+}$, EGTA, urea) from an Amlexanox-Toyopearl column **(C)** and a fluphenazine-Sepharose column **(D)**.

2. Because of the hydrophobic nature of the drugs used here, the drug affinity columns work largely similar to calcium-dependent hydrophobic interaction chromatography. Hence, it is extremely important to perform these experiments at salt concentrations and temperatures as indicated, otherwise results will vary considerably.

## References

1. Jamieson, G. A. and Vanaman, T. C. (1979) Calcium-dependent affinity chromatography of calmodulin on an immobilized phenothiazine. *Biochem. Biophys. Res. Commun.* **90,** 1048–1056.
2. Marshak, D. R., Watterson, D. M., and Van Eldik, L. J. (1981) Calcium-dependent interaction of S100b, troponin C, and calmodulin with an immobilized phenothiazine. *Proc. Natl. Acad. Sci. USA* **78,** 6793–6797.

3. Endo, T., Tanaka, T., Isobe, T., Kasai, H., Okuyama, T., and Hidaka, H. (1981) Calcium-dependent affinity chromatography of S-100 and calmodulin on calmodulin antagonist-coupled Sepharose. *J. Biol. Chem.* **256,** 12,485–12,489.
4. Vogel, H. J, Lindahl, L., and Thulin, E. (1983) Calcium-dependent hydrophobic interaction chromatography of calmodulin and related calcium-binding proteins. *FEBS Lett.* **157,** 241–246.
5. Williams, A. and Ibrahim, I. A. (1981) A mechanism involving cyclic tau-tomers for the reaction with nucleophiles of the water-soluble peptide coupling reagent 1-ethyl-3-(dimethyl amminopropyl) carbodiimide (EDC). *J. Am. Chem. Soc.* **103,** 7090–7095.
6. Gilles, M. A., Hudson, A. Q., and Borders, C. L. (1990) Stability of water-soluble carbodiimides in aqueous solution. *Anal. Biochem.* **184,** 244–248.
7. Sundberg, L. and Porath, J. (1974) Preparation of adsorbents for biospecific affinity chromatography. Attachment of group-containing ligands to insoluble polymers by means of bifuctional oxiranes. *J. Chromatogr.* **90,** 87–98.
8. Silvanovich, M. P. and Hill, R. D. (1976) Affinity chromatography of cereal alpha-amylase. *Anal. Biochem.* **73,** 430–433.
9. Uy, R. and Wold, F. (1977) 1,4-Butanediol diglycidyl ether coupling of carbohydrates to Sepharose: affinity adsorbents for lectins and glycosidases. *Anal. Biochem.* **81,** 98–107.
10. Hermanson, G. T., Mallia, A. K., and Smith, P. K. (1992) *Immobilized Affinity Ligand Tschniques*, (Product #22230), Academic, California.
11. Dean, P. D. G., Johnson, W. S., and Middle, F. A., eds. (1985) *Affinity Chromatography: A Practical Approach*. IRL, Oxford.
12. Oyama, Y., Shishibori, T., Yamashita, K., Naya, T., Nakagiri, S., Maeta, H., and Kobayashi, R. (1997) Two distinct anti-allergic drugs, amlexanox and cromolyn, bind to the same kinds of calcium binding proteins, except calmodulin, in bovine lung extract. *Biochem. Biophys. Res. Commun.* **240,** 341–347.
13. Shishibori, T., Oyama, Y., Matsushita, O., Yamashita, K., Furuichi, H., Okabe, A., et al. (1999) Three distinct anti-allergic drugs, amlexanox, cromolyn and tranilast, bind to S100A12 and S100A13 of the S100 protein family. *Biochem. J.* **338,** 583–589.
14. Yamashita, K., Oyama, Y., Shishibori, T., Matsushita, O., Okabe, A., and Kobayashi, R. (1999) Purification of bovine S100A12 from recombinant *Escherichia coli*. *Protein. Expr. Purif.* **16,** 47–52.
15. Schägger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166,** 368–379.
16. Landar, A., Rustandi, R. R., Weber, D. J., and Zimmer, D. B. (1998) S100A1 utilizes different mechanisms for interacting with calcium-dependent and calcium-independent target proteins. *Biochemistry* **37,** 17,429–17,438.
17. Osterloh, D., Ivanenkov, V. V., and Gerke, V. (1998) Hydrophobic residues in the *C*-terminal region of S100A1 are essential for target protein binding but not for dimerization. *Cell Calcium* **24,** 137–151.

# 26

# Enzymatic Assays to Compare Calmodulin Isoforms, Mutants, and Chimeras

**Michael P. Walsh, Jacquelyn E. Van Lierop, Cindy Sutherland, Ritsu Kondo, and J. David Johnson**

## 1. Introduction

Calmodulin (CaM), the principal protein mediator of cellular $Ca^{2+}$ signals, interacts with some 80 target proteins, many of which are enzymes that are activated by CaM in a $Ca^{2+}$-dependent manner. Mammalian genomes contain at least three differentially regulated CaM genes that encode the same protein ***(1)***. On the other hand, multiple genes encode several CaM isoforms in plants. For example, the soybean genome contains at least five CaM genes that encode four distinct isoforms ***(2)***.

Studies of CaM chimeras, mutants, and isoforms indicate that the interactions of CaM with different target proteins and the molecular mechanisms of activation of CaM-dependent enzymes vary depending on the target enzyme. Thus, for example, the soybean CaM isoform SCaM-1 (90.5% identical in sequence to human CaM) activates calcineurin (CaN; type 2B protein serine/threonine phosphatase) but SCaM-4 (77% sequence identity to human CaM) does not; on the other hand, SCaM-4 activates nitric oxide synthase (NOS) but SCaM-1 does not ***(3)***. The fact that SCaM-4 acts as a competitive inhibitor of SCaM-1-mediated activation of CaN and SCaM-1 acts as a competitive inhibitor of SCaM-4-mediated activation of NOS indicates that both plant isoforms bind to the same site on the target enzyme, but in one case (different for each target enzyme) binding is not coupled to activation. Several instances have been described of site-specific mutations in CaM resulting in loss of activation of a specific target enzyme with little effect on binding. For example, replacement of M144 by V in mammalian CaM has no effect on activation of $Ca^{2+}$/CaM-dependent cyclic nucleotide 3':5'-phosphodiesterase (PDE) or CaN, but con-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

verts CaM from an activator to a competitive antagonist of NOS ***(4)***. A great deal can be learned, therefore, from studying the effect of CaM mutants and isoforms on the activation of various CaM target enzymes.

In this chapter, we describe assay methods for such a comparison using five CaM target enzymes: PDE, CaN, NOS, myosin light chain kinase (MLCK), and $Ca^{2+}$/CaM-dependent protein kinase II (CaM kinase II). CaM-dependent PDE catalyzes the hydrolysis of cAMP and cGMP to the corresponding 5'-nucleoside monophosphates, thereby terminating cyclic nucleotide signaling ***(5)***. This enzyme, therefore, represents an important point of cross-talk between $Ca^{2+}$ and cyclic nucleotide signaling pathways. CaN is a $Ca^{2+}$/CaM-dependent protein serine/threonine phosphatase with a relatively narrow substrate specificity ***(6)***. It has diverse regulatory roles, e.g., T-lymphocyte activation, regulation of neurotransmitter release, and modulation of long-term changes in synaptic plasticity. It is the target of the immunosuppressive drugs, FK506 and cyclosporin A. NOS catalyzes the formation of the intercellular messenger nitric oxide (NO) from L-arginine. NO is a major regulator in the nervous, immune, and cardiovascular systems ***(7)***. There are two classes of NOS: constitutive and inducible. Constitutive NOS is regulated by $Ca^{2+}$-dependent interaction with CaM, whereas inducible NOS contains tightly bound CaM that is not dissociated by chelation of $Ca^{2+}$ ions. MLCK plays a key role in the regulation of smooth muscle contraction and nonmuscle motility via the specific phosphorylation of myosin II ***(8)***. Finally, CaM kinase II, unlike MLCK, has a large number of substrates and is, therefore, involved in the regulation of diverse physiological processes including synaptic transmission, secretion, and gene expression ***(9)***.

For activation of PDE, CaN, and NOS by CaM, we describe continuous assays that allow enzymatic activity to be monitored by changes in fluorescence or absorption upon enzyme activation. For activation of MLCK, CaM kinase II, and NOS by CaM, we describe radioisotope-based assays that follow incorporation of $^{32}P$ from [$\gamma$-$^{32}P$]ATP into the 20-kDa light chain of myosin II ($LC_{20}$), incorporation of $^{32}P$ from [$\gamma$-$^{32}P$]ATP into caldesmon and conversion of L-[$^{14}C$] arginine to L-citrulline and NO, respectively.

For most of these enzymes, CaM binds and removes a pseudosubstrate (autoinhibitory) domain from the enzyme's active site resulting in enzyme activation. CaM activation of NOS is more complicated ***(10,11)***. CaM binds NOS between its *N*-terminal oxygenase domain and its *C*-terminal reductase domain and it stimulates NADPH oxidation and reduction of bound flavin in NOS's reductase domain ***(11)***. CaM also facilitates electron transfer from the reductase domain to the heme-containing oxygenase domain, resulting in the conversion of L-Arg to NO and L-citrulline. Finally, CaM can stimulate the intermolecular transfer of electrons from NOS's reductase domain to exogenous electron acceptors like cytochrome *c* (cyt *c*). Thus, for activation of NOS

by CaM, we describe four assays that measure these different CaM-dependent catalytic activities: an oxyhemoglobin ($HbO_2$) assay and a citrulline assay that measure NO synthesis occurring in the heme domain ***(12,13)***; an NADPH oxidation assay that measures this reductase domain function that is also dependent on a functional heme domain ***(12,13)***; and a cyt *c* reduction assay that selectively measures the function of NOS's reductase domain ***(14)***.

## 2. Materials

### 2.1. Phosphodiesterase Assay

1. A spectrofluorimeter that can monitor fluorescence intensity as a function of time.
2. Quartz cuvets.
3. Purified CaM or CaM mutant (100–300 μ*M* stocks that are stable for years if frozen).
4. The fluorescent substrate, 2'-methylanthraniloyl-cyclic GMP (Mant-c-GMP). Stable for years if frozen as a 2.5-m*M* stock.
5. 10 m*M* MOPS, 200 μ*M* EGTA, 90 m*M* KCl, 0.5 m*M* $CaCl_2$, 5 m*M* $MgCl_2$ (pH 7.0).
6. Purified PDE (approx 1 mg/mL). Stable for years at –20°C in 50% glycerol.

### 2.2. Calcineurin Assay

1. The spectrofluorimeter, cuvets, and CaM stocks aforementioned.
2. The fluorescent substrate, 4-methylumbelliferyl phosphate (MUF). Stable for years if frozen as a 10-m*M* stock in the absence of contaminating phosphatases.
3. 50 m*M* Tris-HCl, 200 μ*M* EGTA, 0.5 m*M* $CaCl_2$, 5 m*M* $MgCl_2$ (pH 7.4). This buffer should be boiled for 10 min to denature contaminating phosphatase.
4. Purified CaN (0.5–1 mg/mL). Stable for years if frozen in aliquots to avoid freeze-thawing.

### 2.3. NOS Oxyhemoglobin Assay

1. UV/VIS spectrophotometer capable of recording absorption as a function of time. A temperature control device is optional. Disposable 1-mL plastic cuvets can be used.
2. 50 m*M* HEPES buffer (pH 7.5).
3. $HbO_2$ (approx 300 μ*M* stock solution), prepared as a 12.5 mg/mL solution of commercially available >95% pure $HbO_2$ (Sigma) in 50 m*M* HEPES (pH 7.5) and stored at –80°C.
4. Dithiothreitol (DTT) (0.3 *M* stock in double-distilled water [$ddH_2O$]).
5. L-Arg (1 *M* stock in $ddH_2O$).
6. $CaCl_2$ (0.5 *M* stock in $ddH_2O$).
7. NADPH (preweighed 1-mg vials). Stable for months if stored dry at room temperature.
8. Flavin adenine dinucleotide (FAD), flavin mononucleotide (FMN), and tetrahydrobiopterin ($H_4B$) (approx 4 m*M* stock in $ddH_2O$).
9. Catalase (100,000 U/mL stock in $ddH_2O$).
10. Superoxide dismutase (SOD; 10,000 U/mL stock in $ddH_2O$).

11. Bovine serum albumin (BSA; 10 mg/mL stock in $ddH_2O$).
12. Purified CaM (100–300 μ*M* stock).
13. Purified NOS (2–4 mg/mL). Stable for months when frozen at –80°C. Avoid freeze-thawing by storing as small aliquots.
14. All stocks are stable for 6 mo or longer when kept frozen at –20°C, unless otherwise stated.

### 2.4. NOS NADPH Oxidation Assay

Same reagents as for the oxyhemoglobin assay, except $HbO_2$, SOD, and BSA are not required and a more dilute solution of catalase (10,000 U/mL stock in $ddH_2O$) is used.

### 2.5. NOS Cyt c Reduction Assay

1. 50 m*M* HEPES buffer (pH 7.5).
2. $CaCl_2$ (0.5 *M* stock in $ddH_2O$).
3. NADPH (approx 10 mg/mL; 10 m*M* stock solution in $ddH_2O$).
4. Cyt *c* (5 m*M* stock solution in buffer).
5. Purified CaM (100–300 μ*M* stock).
6. All of the above stocks are stable for 6 mo or longer when kept frozen at –20°C.
7. Purified NOS (2–4 mg/mL). Stable for months when frozen at –80°C. Avoid freeze-thawing by storing as small aliquots.

### 2.6. NOS Citrulline Assay

1. Same as for the oxyhemoglobin assay, except a scintillation counter, but no spectrophotometer, $HbO_2$, SOD, catalase, or BSA are required.
2. L-[$^{14}$C]Arg (200 m*M* stock with a specific activity of approx 3 μCi/μmol). Stable for months when refrigerated.
3. A strong cation exchange resin such as Bio-Rad 50W-X8. Must be converted to the sodium form by washing with 5 vol of 1 N NaOH, then neutralized with 5 vol of water.

### 2.7. MLCK Assay

1. A scintillation (beta) counter.
2. A temperature-controlled water bath.
3. Purified CaM or CaM mutant stocks aforementioned.
4. Purified substrate: $LC_{20}$ (the 20 kDa light chain of myosin II). Stable for years at –80°C.
5. 50 m*M* Tris-HCl (pH 7.5), 120 m*M* KCl, 8 m*M* $MgCl_2$, 0.2 m*M* $CaCl_2$, 2 m*M* DTT, 0.2% (v/v) Tween-80.
6. Purified MLCK (0.2 mg/mL). Stable for years at –80°C.
7. [γ-$^{32}$P]ATP (>5000 Ci/mmol). Stock solution of 6 m*M* with a specific activity of 150–200 cpm/pmol.
8. P81 phosphocellulose paper (Whatman).

### 2.8. CaM Kinase II Assay

1. A scintillation counter, water bath, P81 paper and purified CaM as in **Subheading 2.7.**
2. Caldesmon substrate copurified with CaM kinase II (1–2 mg/mL). Stable for years at −80°C.
3. 50 m*M* Tris-HCl (pH 7.5), 20 m*M* $MgCl_2$, 0.4 m*M* $CaCl_2$, 0.2% (v/v) Tween-80.
4. [$\gamma$-$^{32}$P]ATP stock solution of 2 m*M* with a specific activity of approx 300 cpm/pmol.

## 3. Methods

### 3.1. Phosphodiesterase Assay

We have previously shown that CaM-dependent PDE will hydrolyze 2'-methylanthraniloyl-cyclic GMP (Mant-c-GMP) to 3'-Mant-GMP resulting in a twofold decrease in its fluorescence ***(15)***. If Mant-c-GMP fluorescence is followed as a function of time, it provides a continuous assay for PDE activity. **Figure 1** shows the CaM- dependent activation of PDE as monitored by the decrease in Mant-c-GMP fluorescence.

To conduct this assay:

1. Add 8 $\mu$*M* Mant-c-GMP to 11 mL of 10 m*M* MOPS, 200 $\mu$*M* EGTA, 90 m*M* KCl, 0.5 m*M* $CaCl_2$, 5 m*M* $MgCl_2$ (pH 7.0) in a plastic tube. This is enough for 10 assays, using a different [CaM] in each 1-mL assay.
2. Mix the solution by inverting 4–5 times and let it equilibrate to the desired assay temperature.
3. Place 1 mL of the solution in a quartz cuvet and place it in the fluorimeter.
4. Set the excitation wavelength to 280 nm and the emission wavelength to 450 nm.
5. Adjust the sensitivity so that the Mant-c-GMP emission is reading approx 80–90% of full scale.
6. Record the fluorescence intensity continuously for 1–2 min, being sure the fluorescence intensity is stable (as in **Fig. 1**).
7. Add approx 15 n*M* of purified PDE and mix rapidly (*see* **Note 1**). Record the fluorescence intensity with time for approx 2 min. The fluorescence should begin to decrease slowly because of basal PDE activity (as shown in **Fig. 1**).
8. Add CaM or CaM mutant at the desired concentration. Mix rapidly and continue following the time-dependent decrease in Mant-c-GMP fluorescence. **Figure 1** shows the rate of decrease in Mant-c-GMP fluorescence when 0, 5, 7.5, 10, 15, or 50 n*M* CaM were added to 1 mL of the buffer + PDE solution aforementioned. As [CaM] increases, the reaction rate increases until saturation occurs.
9. The rate of decrease in Mant-c-GMP fluorescence, at any [CaM], gives the extent of PDE activation. This CaM-dependent activation of PDE can be expressed as fold activation by dividing the rate of the fluorescence decrease in the presence of CaM by the rate in the absence of CaM. In **Fig. 1**, 50 n*M* CaM produces a 50-fold activation of PDE. Alternatively, the rate in the presence of maximal [CaM] can be defined as 100% activation and the rate in the absence of CaM as

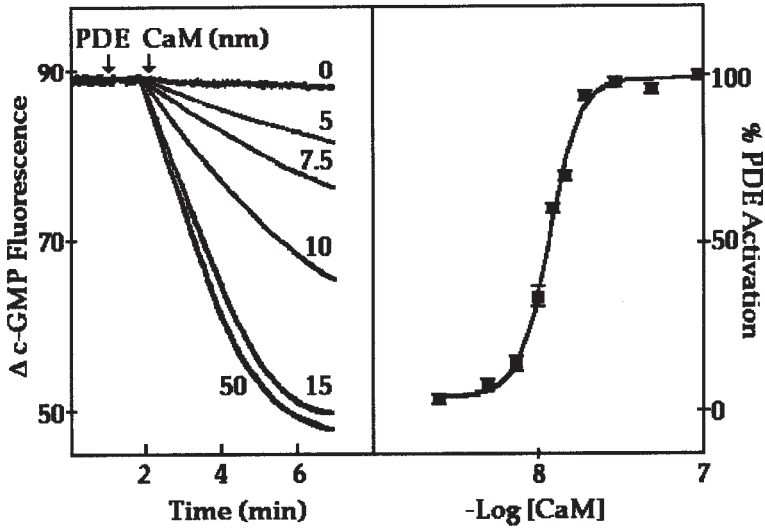


Fig. 1. CaM-dependent activation of PDE-catalyzed hydrolysis of Mant-c-GMP. Additions of PDE and CaM were made as indicated and the experimental conditions are described in **Subheading 3.** The rate of PDE-induced decrease in Mant-c-GMP fluorescence is shown as a function of increasing concentrations of CaM (0, 5, 7.5, 10, 15, and 50 n*M*). A plot of% PDE activation as a function of [CaM] is also shown. 100% activation occurred at approx 20 n*M* CaM and represented a 50-fold increase in the rate of hydrolysis relative to the basal rate.

0% activation. This allows one to plot the% PDE activation as a function of increasing [CaM] as shown in **Fig. 1**.

10. Using this assay, the $K_{act}$ and $V_{max}$ of any CaM mutant or isoform for PDE activation can be quickly and accurately determined (*see* **Notes 2** and **3**). In addition, the effect of CaM inhibitors can be readily tested by determining their ability to inhibit CaM stimulation of PDE (*see* **Note 2**).

### 3.2. Calcineurin Assay

When the protein phosphatase CaN is activated by CaM, it dephosphorylates 4-methylumbelliferyl phosphate (MUF) producing a large time-dependent increase in fluorescence. Anthony et al. ***(16)*** have used MUF to develop a continuous assay for CaN. **Figure 2** shows an example of this assay and CaM stimulation of CaN's dephosphorylation of MUF.

To conduct this assay:

1. Add 100 μ*M* MUF to 11 mL of 50 m*M* Tris-HCl, 200 μ*M* EGTA, 0.5 m*M* $CaCl_2$, 5 m*M* $MgCl_2$ (pH 7.4) in a plastic tube. This is enough for 10 assays, using different [CaM] in each 1-mL assay.

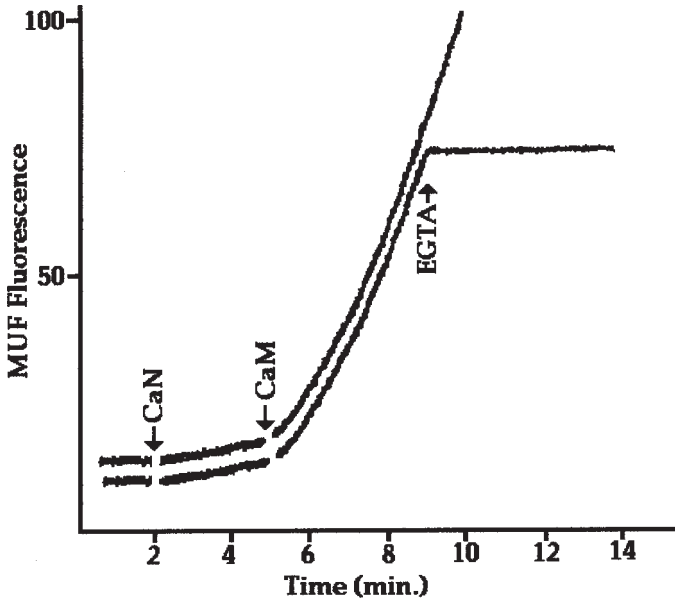


Fig. 2. CaM-dependent activation of CaN-catalyzed dephosphorylation of MUF. CaN, CaM, and EGTA were added as indicated and the experimental conditions are described in **Subheading 3.** CaM stimulated the CaN-induced increase in MUF fluorescence 21-fold.

2. Mix the solution well by inverting 4–5 times and let it equilibrate to the desired assay temperature.
3. Place 1 mL of the solution in a quartz cuvet and place it in the fluorimeter.
4. Set the excitation wavelength to 365 nm and the emission wavelength to 450 nm.
5. Adjust the sensitivity so that the MUF emission is reading approx 5–10% of full scale. When CaN and CaM are added MUF fluorescence will rapidly increase (as in **Fig. 2**) and beginning at low fluorescence will keep it on scale longer.
6. Record the fluorescence intensity continuously for 1–2 min, being sure the fluorescence is stable (as shown in **Fig. 2**).
7. Add approx 15 n*M* of purified CaN, mix rapidly, and record the fluorescence intensity for approx 2 min. The fluorescence should begin to increase slowly because of basal CaN activity (as shown in **Fig. 2**).
8. Add CaM (or CaM mutant) at the desired concentration. Mix rapidly and record the increase in MUF fluorescence as an index of CaN activity. **Figure 2** shows the rate of increase in MUF fluorescence when 14 n*M* CaM was added to 1 mL of the buffer + CaN solution aforementioned. CaM stimulated CaN's activity by approx 20-fold over basal. **Figure 2** also shows that this reaction can be rapidly stopped by the addition of 2 m*M* EGTA, which dissociates CaM from CaN resulting in enzyme inactivation.

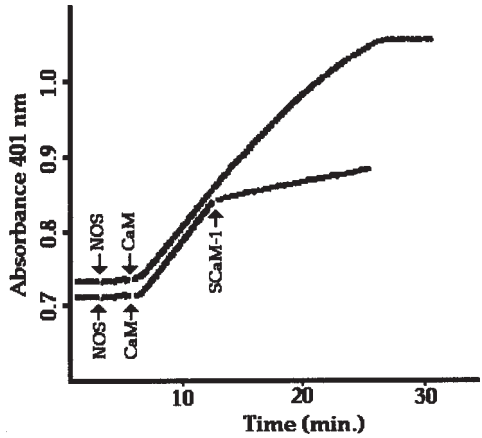


Fig. 3. CaM-dependent activation of NOS-catalyzed NO production as monitored by the oxyhemoglobin assay. NOS and CaM were added at the indicated times and the experimental conditions are described in **Subheading 3.** CaM stimulated NO production approx 10-fold over the basal level. In the second assay shown, NOS was activated by CaM and, during the linear portion of the increase in oxyhemoglobin absorption at 401 nm, SCaM-1 (1 μ*M*) was added. We have previously shown that SCaM-1 is a selective competitive antagonist of NOS *(3)*.

9. The rate of increase in MUF fluorescence, at any [CaM], gives the extent (% or fold) of CaN activation. Using this assay, the effect of any CaM mutant or isoform on CaN activation can be quickly and accurately determined (*see* **Note 2**). In addition, the effect of CaM inhibitors can be readily tested by determining their ability to inhibit CaM stimulation of CaN as shown above.

### 3.3. NOS Assay: Oxyhemoglobin

When CaM activates NOS it produces NO. NO can then rapidly convert oxyhemoglobin to metHb, producing an increase in absorption at 401 nm ($A_{401}$). Changes in $A_{401}$ provide a convenient continuous assay for measurement of CaM stimulation of NOS. An example is shown in **Fig. 3**.

To conduct this assay:

1. Make up 11 mL of 50 m*M* HEPES (pH 7.5), 0.3 m*M* DTT, 1 m*M* L-Arg, 1 m*M* $CaCl_2$, 4 μ*M* FAD, FMN, $H_4B$, 100 U/mL catalase, 10 U/mL SOD, 0.1 mg/mL BSA.
2. Pour this solution into a preweighed 1 mg vial of NADPH. Mix by inverting five times and preincubate at the desired reaction temperature (25–37°C) (*see* **Note 4**).
3. Pipet 1 mL (less the volume of NOS and $HbO_2$ that will be added later) into a cuvet and place the cuvet in the spectrophotometer.
4. Add 3 μ*M* $HbO_2$, mix (*see* **Note 1**) and follow the absorption at 401 nm ($A_{401}$) for 2 min. This baseline should be a stable flat line as shown in **Fig. 3**.

5. Add 20 n*M* NOS (*see* **Notes 5** and **6**), mix and continue to record $A_{401}$. The absorption may increase slightly due to the low basal activity of NOS, as shown in **Fig. 3** (*see* also **Note 5**).
6. Add the desired [CaM], mix, and continue to record $A_{401}$. Upon addition of CaM, NOS will be activated and, as it produces NO, $A_{401}$ will increase. **Figure 3** shows an approx 10-fold enhancement in the rate of increase of $A_{401}$ upon the addition of 50 n*M* CaM.
7. Inhibitors of CaM or NOS can be added during the linear phase of the increase in absorption and inhibition can be followed by their ability to prevent further increases in $A_{401}$ (*see* **Note 2**). This is demonstrated in **Fig. 3**, where 1 μ*M* of a CaM isoform (SCaM-1), which is a competitive antagonist of NOS ***(3)***, was added to inhibit NOS.
8. NOS's enzymatic activity can be calculated from the rate of change in $A_{401}$ using $\Delta\varepsilon$ = 38 m*M*/cm.

### 3.4. NOS Oxidation of NADPH

NADPH undergoes a decrease in absorption at 340 nm ($A_{340}$) upon oxidation and this provides a convenient method for following this function of NOS's reductase domain. In the absence of exogenous electron acceptors (cyt *c* or FeCN), the rate of NADPH oxidation is also dependent upon the presence of a functional electron acceptor (heme) in NOS's oxidase domain (*see* **Note 7**).

To conduct these assays:

1. Make up 11 mL of 50 m*M* HEPES (pH 7.5), 3 m*M* DTT, 1 m*M* L-Arg, 1 m*M* $CaCl_2$, 4 μ*M* FAD, FMN, $H_4B$, and 110 U catalase.
2. Pour the prepared solution into a preweighed 1 mg vial of NADPH. Mix by inverting five times and preincubate at the desired reaction temperature (*see* **Note 4**).
3. Pipet approx 1 mL (less the volume of NOS to be added later) of the above mixture into a cuvet, and follow $A_{340}$ for 2 min in a UV/VIS spectrophotometer. This may show a slight linear decrease if there is basal activity.
4. Add 20 n*M* NOS, mix rapidly and continue to record $A_{340}$. This absorption may decrease slightly because of the basal activity of NOS.
5. Add the desired [CaM], mix rapidly and continue to record $A_{340}$. Upon addition of CaM, NOS will be activated and its rate of NADPH oxidation should increase resulting in a faster decrease in $A_{340}$. Depending on the purity of the NOS, you can expect a 5–100-fold increase in the rate of NADPH oxidation upon addition of CaM.
6. Inhibitors of CaM or NOS can be added during the linear phase of the increase in $A_{340}$ and their inhibition can be followed by their ability to prevent further decreases in absorbance (*see* **Note 2**).
7. Enzyme activation can be calculated from the rate of decrease in $A_{340}$ using $\Delta\varepsilon$ = 6.22 m*M*/cm.
8. It is also possible to conduct the NADPH oxidation assay in the presence of an exogenous electron acceptor like cyt *c* (*see* **Note 7**).

### 3.5. NOS Reduction of Cyt c

Activated NOS can transfer electrons from its reductase domain directly to exogenous electron acceptors like cyt *c*. The increase in cyt *c* absorbance at 550 nm provides a continuous assay of this reductase domain function of NOS.

To conduct this assay:

1. Make up 11 mL of 50 m*M* HEPES (pH 7.5) and 1 m*M* $CaCl_2$ and let it equilibrate to the desired temperature.
2. Pipet approx 1 mL of this solution (less the volume of cyt c and NOS to be added later) into a cuvet and add cyt *c* (50 μ*M*) and NOS (10–20 nM).
3. Mix, then place the cuvet in the spectrophotometer, and follow $A_{550}$ for 2 min. This baseline should be a stable flat line.
4. Add NADPH (100 μ*M*), mix rapidly, and continue to follow $A_{550}$. Upon addition of NADPH, the absorption should show a linear increase (*see* **Note 8**).
5. Quickly add the desired [CaM]. Rapidly mix and continue to follow $A_{550}$. After the CaM addition the rate of increase in $A_{550}$ should accelerate. Under the conditions described, the reaction will remain linear for only 4–5 min. Depending on the purity of the NOS you can expect a 5–25-fold increase in rate compared to basal.
6. Enzymatic activity can be calculated from the rate of increase in $A_{550}$ using $\Delta\varepsilon$ = 21 m*M*/cm.

### 3.6. NO Production as Measured by the Citrulline Assay

This assay is based on the quantitative determination of radioactive citrulline formed by NOS from L-[$^{14}$C]Arg. Although the L-Arg substrate binds to the cation exchange resin (because of its positively charged guanidino moiety), radiolabeled L-citrulline will flow through.

To conduct this assay:

1. Make up 11 mL of 50 m*M* HEPES (pH 7.5) and let it equilibrate to the desired temperature.
2. Add 200 μ*M* L-[$^{14}$C]Arg, 1 m*M* $CaCl_2$ , 4 μ*M* FAD, FMN, $H_4B$.
3. Pour the prepared solution into a preweighed 1 mg vial of NADPH and mix by inverting five times.
4. Pipet approx 300 μL (less the volume of NOS and CaM to be added later) of the aforementioned mixture into plastic tubes.
5. Add the desired [CaM] to each tube. Vortex and preincubate at 37°C for 5 min.
6. Start the reaction by adding 10–25 n*M* NOS. Vortex each tube quickly and incubate at 37°C for 5–20 min, depending on the time over which the reaction is linear (*see* **Note 2**).
7. After incubation for a precise time, stop the reaction by adding 5 μL of 6 *N* trichloroacetic acid to each tube. Vortex and place on ice.
8. Neutralize the pH by adding 250 μL of 1.5 *M* HEPES (pH 7.5) to each tube.
9. Apply the contents of each tube to a separate 1 mL cation exchange column (sodium form).

10. Wash each column with three consecutive 1-mL aliquots of water and collect the flowthrough in 15-mL scintiverse cocktails.
11. Count the cocktails in a scintillation counter (*see* **Note 10**).
12. Determine enzyme activity from the specific activity of the radioactive substrate.

### 3.7. MLCK Assay

MLCK activity is absolutely dependent on $Ca^{2+}$ and CaM. This assay measures the incorporation of $^{32}P$ from [$\gamma$-$^{32}P$]ATP into $LC_{20}$ at a fixed time following addition of radiolabeled ATP to start the reaction. **Figure 4** shows the concentration dependence of CaM activation of MLCK.

To conduct this assay:

1. To 15 μL of buffer in a plastic 1.5-mL microfuge tube (without cap) on ice, add $LC_{20}$ (10 μ*M*), CaM (0–1 μ*M*), and MLCK (0.05 μg/mL).
2. Add distilled, deionized $H_2O$ to give a total volume of 29 μL.
3. Mix the solution well by vortexing and let it equilibrate to 30°C in the water bath.
4. Start the reaction by adding 1 μL of stock [$\gamma$-$^{32}P$]ATP.
5. Incubate the reaction mixture at 30°C for 10 min.
6. With a pipetman, transfer 20 μL of the reaction mixture to a square (1 × 1 cm) of P81 paper and immerse immediately in a glass 600-mL beaker containing 500 mL of 0.5% (v/v) $H_3PO_4$ and a stainless steel wire mesh basket. This stops the reaction.
7. Wash the paper squares three times for 5 min each with stirring in 500 mL of 0.5% (v/v) $H_3PO_4$ and once for 2 min with acetone.
8. Remove the wire basket from the beaker, place on a paper towel and dry the P81 papers with a hair dryer.
9. Transfer the dried paper squares to plastic scintillation vials and quantify $^{32}P$ by Cerenkov counting (no scintillant or other liquid) in a scintillation counter using $^{3}H$ window settings.
10. Determine enzyme activity from the specific activity of the radiolabeled ATP.

### 3.8. CaM Kinase II Assay

This assay measures the incorporation of $^{32}P$ from [$\gamma$-$^{32}P$]ATP into caldesmon at a fixed time following addition of radiolabeled ATP to start the reaction. **Figure 5** shows the characterization of the CaM kinase II/caldesmon preparation: caldesmon phosphorylation is observed only in the presence of both $Ca^{2+}$ and CaM, and is completely inhibited by [Ala9]autocamtide 2, a CaM kinase II inhibitor. **Figure 6** shows the concentration dependence of CaM activation of CaM kinase II.

To conduct this assay:

1. To 15 μL of buffer in a plastic 1.5-mL microfuge tube (without cap) on ice, add CaM (0–10 μ*M*) and caldesmon containing CaM kinase II (0.2 mg/mL).
2. Add distilled, deionized $H_2O$ to give a total volume of 27 μL.

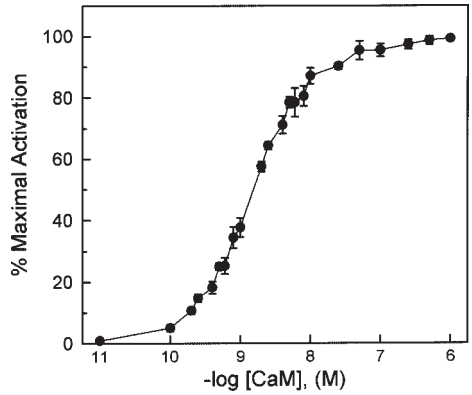


Fig. 4. CaM-dependent activation of MLCK. MLCK activity was measured at the indicated concentrations of bovine brain CaM. Values represent the mean ± SEM ($n$ = 3 – 7, each done in duplicate or triplicate). Maximal activity corresponds to 12.7 μmol $P_i$ incorporated/min.mg MLCK.

3. Mix the solution well by vortexing and let it equilibrate to 30°C in a water bath.
4. Start the reaction by adding 3 μL of stock [$\gamma$-$^{32}$P]ATP.
5. Continue from #5 of MLCK assay (*see* **Subheading 3.7.**).

## 4. Notes

### 4.1. General Considerations

1. For any assay, it is important that the solution be mixed thoroughly after each addition. For the continuous assays, this is achieved by placing Parafilm over the cuvet and mixing the solution before placing it in the fluorimeter or spectrophotometer. Subsequent additions to the cuvet are generally mixed by placing a pipet tip in the cuvet and rapidly drawing 150 μL of solution in and out of the pipet tip five times. Alternatively, a stirring stick or a magnetic stir bar can be used. For the discontinuous assays vortexing is used to assure homogeneity after each addition.
2. For any enzyme assay, it is essential to determine the time over which the reaction is linear. For the continuous assays, this can be done by simply following the change in fluorescence or absorption as a function of time. It is imperative that any CaM dose- response curves or inhibitor studies be conducted while the change in signal is in the linear phase. Depending on the time of linearity, this may allow determination of the effect of one or many [CaM] per assay. For example, the continuous PDE assay is so rapid (over in 3–4 min) that we add one [CaM] per assay. The continuous CaN assay remains linear for 25 min, even with maximal [CaM], and it is, therefore, possible to determine the effect of several [CaM] in one assay. If the CaM dose-response curve allows you to use multiple

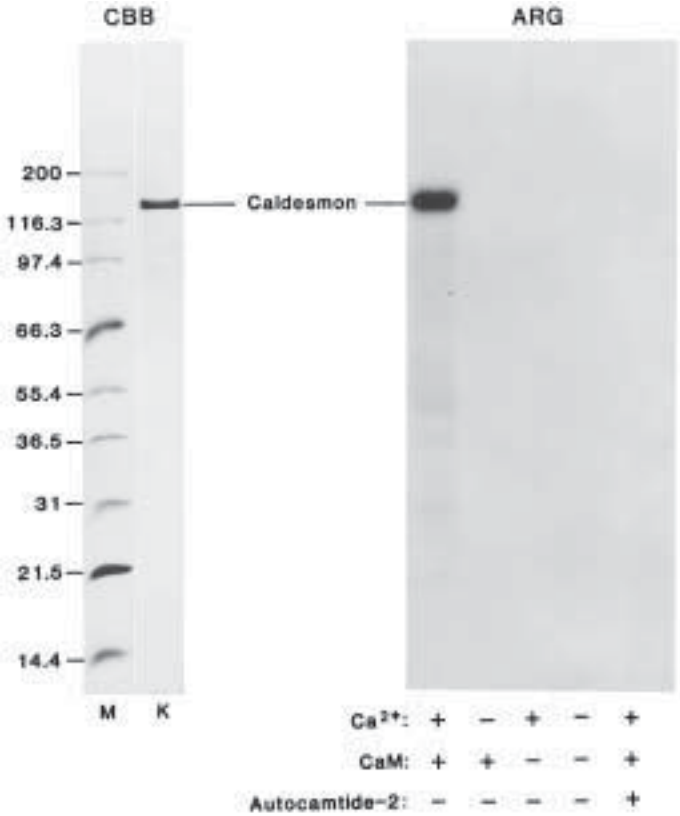


Fig. 5. Characterization of the CaM kinase II/caldesmon preparation. Caldesmon containing CaM kinase II was incubated at 30°C for 10 min with [γ-$^{32}$P]ATP under the indicated conditions. Reactions were stopped by addition of an equal volume of SDS gel sample buffer (20 μL) and boiling. Samples were subjected to SDS-PAGE and autoradiography. CBB: Coomassie Blue-stained SDS gels of $M_r$ markers (*M*) and the CaM kinase II/caldesmon preparation (*K*). ARG: autoradiograph showing the $Ca^{2+}$- and CaM-dependence of caldesmon phosphorylation and inhibition by the CaM kinase II inhibitor peptide, [Ala9]autocamtide 2 (10 μ*M*). Where present, [CaM] was 1 μ*M*.

[CaM] per assay, it is recommended that at least one identical [CaM] point be included in each assay. This provides a check for internal consistency.

For the noncontinuous assays, linearity must be determined by stopping the reaction at various points in time and quantifying product formation or substrate loss. The time over which the reaction is linear depends on the concentration of enzyme and substrate used and must be verified before selecting a suitable time (on the linear portion of the curve relating product formation (or substrate loss) and time).

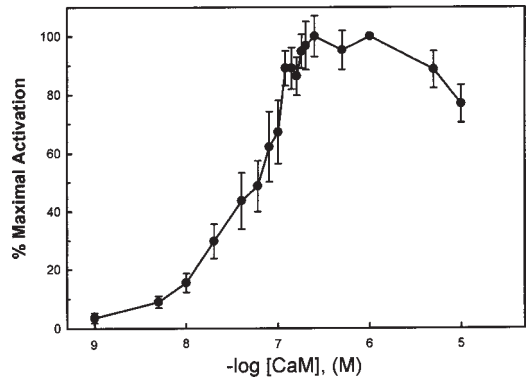


Fig. 6. CaM-dependent activation of CaM kinase II. CaM kinase II activity was measured at the indicated concentrations of bovine brain CaM. Values represent the mean ± SEM ($n = 4$, each done in triplicate). Maximal activity corresponds to 0.7 nmol $P_i$ incorporated/min.mg caldesmon substrate.

For CaM inhibition assays, we first add CaM, then inhibitor and finally start the reaction by addition of the enzyme. In the continuous assays, CaM inhibitors can be added after the reaction has been started as shown in **Figs. 2** and **3**. If this is done, it is important that the reaction rate is linear for a sufficient period of time to allow disruption of the active CaM-enzyme complex.

3. If these assays are being used to study the CaM-dependent activation of enzymes, it is essential that the purified target enzymes contain minimal amounts of CaM. If this is not the case, the basal activity will be high before addition of CaM and this basal rate can be inhibited by the addition of CaM inhibitors or $Ca^{2+}$ chelators such as EGTA. Removal of CaM is not currently possible for inducible NOS.

### 4.2. NOS Assays

4. NOS activities are optimal at pH 7.5 and 37°C and are very pH and temperature sensitive. These parameters must be held constant during a reaction.
5. The amount of NOS used depends on its purity. Typically, CaM produces maximal activation at a 1:1 molar ratio. By using lower [NOS] the reaction will remain linear for a longer time. Because all the NOS assays described here require soluble enzyme, if eNOS is used it must first be solubilized.
6. SOD and catalase are included in the oxyhemoglobin assay to prevent superoxide (formed by uncoupled NADPH oxidation) from reacting with NO, and to prevent $H_2O_2$ (formed by SOD) from converting $HbO_2$ to metHb and other higher oxidation states, respectively.
7. It is possible to conduct the NADPH oxidation assay in the presence of an exogenous electron acceptor such as cyt *c*. In the presence of cyt *c*, NADPH oxidation is no longer dependent on a functional oxygenase domain in NOS. NADPH oxi-

dation then selectively reports the activation of NOS's reductase domain. When cyt *c* is included NADPH oxidation should be measured at 337.5 nm to eliminate interference from cyt *c* absorption ***(11)***.

8. The cyt *c* reduction assay should be initiated by addition of NADPH. Under these conditions, the reaction occurs rapidly and is linear for only approx 4 min even before the addition of CaM. Therefore, CaM must be added quickly to avoid depletion of NADPH.
9. The cyt *c* reduction assay requires 5–10 times less NOS than the other assays. If too much NOS is used, the reaction proceeds too quickly and is linear for too short a time to follow activation by CaM. Generally, using less NOS allows the reaction to be linear for several minutes and this allows determination of the effect of one [CaM] per assay.
10. In the citrulline assay each sample must be corrected for the amount of radioactive L-Arg that flows through the column and is collected with the L-citrulline recovery. This is accomplished by subtracting the radioactive counts of a sample that does not contain NADPH (or subtracting the counts of a reaction which was stopped at zero time) from all other samples. Citrulline recovery can be determined with [$^3$H]citrulline (approx 30 μ*M*) under identical conditions. Furthermore, the citrulline assay can only be used if any contaminating L-Arg in the NOS sample has been removed by desalting.

### 4.3. MLCK Assays

11. Alternative substrates to $LC_{20}$ can be used in this assay: purified myosin II or a synthetic peptide corresponding, for example, to residues 11–23 of $LC_{20}$ (with the sequence KKRPQRATSNVFA), i.e., containing ser19, the site of phosphorylation by MLCK. If myosin II is used, it is necessary to ensure that the preparation is not contaminated by CaM, MLCK, or myosin light chain phosphatase. The CaM activation curve is shifted to the right if the synthetic peptide substrate is used rather than $LC_{20}$ or myosin II.
12. Stock MLCK should be diluted just prior to its addition to the assay mixture. Loss of activity is observed if MLCK is stored in dilute solution.

### 4.4. CaM Kinase II Assays

13. We copurify CaM kinase II with its substrate caldesmon since the stability of the enzyme is much greater than it is following their separation.

## References

1. Lu, K. P. and Means, A. R. (1993) Regulation of the cell cycle by calcium and calmodulin. *Endocr. Rev.* **14,** 40–58.
2. Lee, S. H., Kim, J. C., Lee, M. S., Heo, W. D., Seo, H. Y., Yoon, H. W., et al. (1995) Identification of a novel divergent calmodulin isoform from soybean which has differential ability to activate calmodulin-dependent enzymes. *J. Biol. Chem.* **270,** 21,806–21,812.

3. Cho, M. J., Vaghy, P. L., Kondo, R., Lee, S. H., Davis, J. P., Rehl, R., et al. (1998) Reciprocal regulation of mammalian nitric oxide synthase and calcineurin by plant calmodulin isoforms. *Biochemistry* **37,** 15,593–15,597.
4. Kondo, R., Cho, M. J., and Johnson, J. D. (1999) A point mutation in a plant calmodulin is responsible for its inhibition of nitric oxide synthase. *FASEB J.* **13,** A1532.
5. Zhao, A. Z., Yan, C., Sonnenburg, W. K., and Beavo, J. A. (1997) Recent advances in the study of $Ca^{2+}$/CaM-activated phosphodiesterases: expression and physiological functions. *Adv. Second Messenger Phosphoprotein Res.* **31,** 237–251.
6. Klee, C. B., Ren, H., and Wang, X. (1998) Regulation of the calmodulin-stimulated protein phosphatase, calcineurin. *J. Biol. Chem.* **273,** 13,367–13,370.
7. Knowles, R. G. and Moncada, S. (1994) Nitric oxide synthases in mammals. *Biochem. J.* **298,** 249–258.
8. Gallagher, P. J., Herring, B. P., and Stull, J. T. (1997) Myosin light chain kinases. *J. Muscle Res. Cell Motil.* **18,** 1–16.
9. Braun, A. P. and Schulman, H. (1995) The multifunctional calcium/calmodulin-dependent protein kinase: from form to function. *Annu. Rev. Physiol.* **57,** 417–445.
10. Gachhui, R., Abu-Soud, H. M., Ghosha, D. K., Presta, A., Blazing, M. A., Mayer, B., et al. (1998) Neuronal nitric-oxide synthase interaction with calmodulin-troponin C chimeras. *J. Biol. Chem.* **273,** 5451–5454.
11. Stevens-Truss, R., Beckingham, K., and Marletta, M. A. (1997) Calcium binding sites of calmodulin and electron transfer by neuronal nitric oxide synthase. *Biochemistry* **36,** 12,337–12,345.
12. Hevel, J. M. and Marletta, M. A. (1994) Nitric oxide synthase assays. *Methods Enzymol.* **233,** 250–258.
13. Stuehr, D. J. and Griffith, O. W. (1996) Purification, assay and properties of mammalian nitric oxide synthases, in *Methods in Nitric Oxide Research* (Feelisch, M. and Stamler, J. S., eds.), Wiley, New York, pp. 177–186.
14. Richards, M. K., Clague, M. J., and Marletta, M. A. (1996) Characterization of C415 mutants of neuronal nitric oxide synthase. *Biochemistry* **35,** 7772–7780.
15. Johnson, J. D., Walters, J. D., and Mills, J. S. (1987) A continuous fluorescence assay for cyclic nucleotide phosphodiesterase hydrolysis of cyclic GMP. *Anal. Biochem.* **162,** 291–295.
16. Anthony, F. A., Merat, D. L., and Cheung, W. Y. (1986) A spectrofluorimetric assay of calmodulin-dependent protein phosphatase using 4-methylumbelliferyl phosphate. *Anal. Biochem.* **155,** 103–107.

# 27

# Gene Expression in Transfected Cells

**Kate Hughes, Juha Saarikettu, and Thomas Grundström**

## 1. Introduction

A general approach to address the biological function of a calcium-binding protein, or another protein, in living cells is to increase or decrease the activity of the protein in the cell and analyze the effects on cell functions. In many cases, it is desirable to determine the effects of overexpressing the protein or a constitutively active or dominantly negative derivative, or to express the protein in a cell that normally lacks it. This is achieved by introducing its gene exogenously. The cDNA for the protein is cloned downstream of an active promoter in a plasmid designed for expression in mammalian cells. This expression plasmid is then transfected into the cell.

Mammalian cells can be transfected by a number of methods. Among the more common techniques used today are cellular uptake of DNA prompted by chemical means such as cationic liposomes or DEAE-dextran, and the physical method of electroporation. These methods are relatively easy, and result in transfection of a large proportion of the cell population. Transfection mediated by chemical means rely on the responsiveness of the cell to the reagent used, and so the efficiency varies dramatically between cell types. Electroporation, however, is successful for a diversity of cell types, including many that are resistant to other methods of transfection. Furthermore, the lack of chemicals in the electroporation procedure reduces the risk of side effects on the cells. Liposome and DEAE-dextran transfection reagents can be purchased as kits with accompanying protocols from commercial sources such as Gene Therapy Systems, Invitrogen, Life Technologies, and Promega, and general overviews of these and other transfection procedures can be found in **refs. *1–3***. In this chapter, we will focus on electroporation.

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

The electroporation procedure entails mixing cells with DNA and subjecting them to a high-voltage electric field, which transiently permeabilises the cell membrane allowing DNA to enter the cell (reviewed in **refs. *4*** and ***5***). Once inside the cell, the DNA is transported to the nucleus and transcription is initiated from the promoter of the expression plasmid. The choice of promoter is important in that it should enable a high level of transcription of the cDNA in the chosen cell type. Viruses such as the human cytomegalovirus (CMV), simian virus 40 (SV40), and Rous sarcoma virus (RSV) have naturally evolved strong promoters that function in a variety of cell types and are often the best choice. The level of protein expression also depends on a number of other factors, such as efficiency of transfection, cell type, and regulation of the protein by post-transcriptional mechanisms. For example, calmodulin is efficiently regulated post-transcriptionally ***(6–8)***, warning that profound increases in mRNA levels by transfection will not necessarily result in correspondingly profound increases in protein levels.

To monitor the success of a transfection, it is common to include an internal control plasmid. In analogy to the expression construct, this plasmid contains the gene for an enzyme cloned downstream of a constitutively active promoter. At an appropriate time after transfection, a sample of cells is lysed and assayed for enzyme activity to ensure that the DNA was successfully delivered into the cells. By comparing the enzyme activity in the same number of cells from different transfections, it is also possible to normalize and thereby compare different transfections.

It is far beyond the scope of this chapter to discuss the many possible analyses of functional consequences of (over-) expressing a calcium-binding protein or a protein regulated by a calcium-binding protein in living cells. However, one general process that can be readily assayed is the effect of the transfected protein on transcription from a gene control region. The regulatory sequence of interest is cloned into a plasmid upstream of a gene for a reporter enzyme. The chosen enzyme should be absent from mammalian cells and its activity easily measured. Luciferase is usually used for this purpose, as it can be measured by a convenient and very sensitive assay, whereas a gene for another luciferase enzyme that can be assayed independently, β-galactosidase or chloramphenicol acetyl transferase (CAT) is used in the internal control plasmid. The amount of reporter enzyme produced, determined by measuring its activity in a cell lysate, is proportional to the amount of transcription initiated from the regulatory sequence. The appeal of this simple method is that it also allows mutational analysis to define regions of a regulatory element that are important for the studied transcriptional regulation, as well as analyses of effects through individual transcription factors by using isolated specific DNA binding sites.

A standard protocol for transient transfection of suspension cells by electroporation and analysis of the expression of reporter genes is described

below. This technique is equally applicable to adherent cells, but requires that the cells are in suspension during the procedure. A number of alternative protocols have been developed to allow electroporation of adherent cells in their attached state ***(9–11)***.

## 2. Materials

### 2.1. Electroporation

1. Basic cell-culture equipment: 37°C carbon dioxide incubator and cell-culture hood.
2. Cell-culture medium (as used to culture your cells).
3. 25-$cm^2$ (50 mL) cell-culture flasks.
4. Sterile 1.5-mL and 50-mL tubes.
5. DNA constructs (*see* **Note 1**):
   a. Expression plasmid with your cDNA inserted downstream of a desired promoter. The parental "empty" plasmid, lacking the cDNA, is needed as a control to ensure that any phenotype seen is caused by the expression of the transfected cDNA.
   b. Internal control plasmid for the normalization of transfections, for example a β-galactosidase gene under the control of a constitutively active promoter.
   c. Luciferase reporter plasmid containing a luciferase gene under the control of the inserted DNA-regulatory element whose activity you want to study.

   Mammalian expression plasmids with different promoters and enzyme reporter genes are available from a number of commercial sources, such as CLONTECH, Invitrogen, Promega, and Stratagene.
6. Sterile electroporation cuvets with an electrode gap of 0.4 cm. These are available from a number of companies, including Bio-Rad, Invitrogen, and Life Technologies (*see* **Note 2**).
7. Electroporation system (pulse generator). These are available from a number of companies, a popular model being the Gene Pulser® II Electroporation System (Bio-Rad).
8. Sterile Pasteur pipets.

### 2.2. Lysis of the Cells

1. 1.5-mL and 15-mL tubes.
2. Phosphate-buffered saline (PBS): 137 m*M* NaCl, 2.7 m*M* KCl, 10 m*M* $Na_2HPO_4$, 1.8 m*M* $KH_2PO_4$. Adjust the pH to 7.4 with HCl. Store at room temperature.
3. Lysis buffer that permits β-galactosidase and luciferase assays (for example, Reporter Lysis Buffer, Promega).

### 2.3. β-Galactosidase and Luciferase Assays

1. 1.5-mL tubes.
2. β-galactosidase assay buffer: 60 m*M* $Na_2HPO_4$, 40 m*M* $NaH_2PO_4$, 10 m*M* KCl, 1 m*M* $MgSO_4$, 50 m*M* β-mercaptoethanol. Adjust the pH to 8.0 with NaOH. The buffer can be stored at room temperature for several months.
3. ONPG (o-nitrophenyl-β-D-galactopyranoside) solution. Prepare prior to use by dissolving in β-galactosidase assay buffer to a concentration of 0.8 mg/mL.

4. 1 *M* $Na_2CO_3$.
5. Spectrophotometer.
6. Luciferase assay kit (for example, Luciferase Assay System, Promega).
7. Luminometer, available from many companies including Turner Designs and PharMingen.
8. Luminometer cuvets.

## 3. Methods

### 3.1. Electroporation

The efficiency of transfection is dependent on the growth phase of the cells. For optimal transfections, grow the cells so that they are in mid-log growth phase the day of transfection. Keep the cells sterile throughout the following procedure.

1. Prewarm the cell-culture medium to 37°C, and for each transfection aliquot 10 mL of prewarmed cell-culture medium to a 25 $cm^2$ (50 mL) cell-culture flask and place it in the 37°C carbon dioxide incubator.
2. For each transfection, aliquot the following DNA solutions in a sterile 1.5-mL tube:
   a. 10 μg of expression plasmid containing the cDNA, or 10 μg of "empty" expression plasmid (without the cDNA) as a control (*see* **Note 3**).
   b. 2 μg of internal control plasmid for normalization of transfections.
   c. 2 μg of luciferase reporter plasmid.
3. Count the cell density of the cell culture. Ten million cells are needed for each transfection.
4. Centrifuge enough cells for your transfections at 250*g* for 10 min at room temperature. Be aware that you might lose cells after centrifugation.
5. Resuspend the cells in prewarmed cell-culture medium (*see* **Note 4**) to an approximate density of 30 million cells/mL. Count the cells again and dilute them in prewarmed medium to a final density of 20 million cells/mL.
6. In the 1.5-mL tubes from **step 2**, mix 0.5 mL of cell suspension (i.e., 10 million cells) with the DNA by gentle pipetting. The cells settle easily, so gently mix the cell suspension before taking each 0.5-mL aliquot. Transfer the cell/DNA mixture to a sterile electroporation cuvet.
7. Preincubate the cuvets for 5 min if necessary (*see* **Note 5**).
8. Set up the electroporation system (pulse generator) by selecting a voltage suitable for the cells (*see* **Note 6**). Adjust the capacitance or length of time of the electric pulse according to the manual of the pulse generator. For example, using the Gene Pulser® II system (Bio-Rad), set the capacitance to 950 μF, resulting in an electric pulse time of 15–20 ms.
9. Be sure to remove all liquid from the outside of the cuvet, for example, by drying with a tissue, before electroporating. Just before inserting the cuvet into the shocking chamber, gently flick it a couple of times to mix the cells. Electroporate the cells.
10. After electroporation allow the cells to recover by letting them stand at room temperature for 5 min (*see* **Note 5**).

11. Using a sterile Pasteur pipet, gently transfer all of the 0.5-mL transfected cells, including a white aggregate of cells that forms as a result of the electroporation, to the cell-culture flask containing 10 mL prewarmed medium from **step 1**. Be sure to transfer all of the cells by rinsing the cuvet with medium.
12. Incubate the transfected cells in the 37°C carbon dioxide incubator for 8–72 h (*see* **Note 7**). During the course of a long incubation, fast-growing cells may have to be diluted to prevent the culture from becoming too dense.

### 3.2. Lysis of the Cells

1. Harvest the 10.5 mL of transfected cells from **Subheading 3.1., step 12** in 15-mL tubes by centrifugation at 250*g* for 10 min at room temperature.
2. Remove the supernatant. Resuspend the cells in 1 mL PBS and transfer them to 1.5-mL tubes. Centrifuge at 250*g* for 10 min at room temperature.
3. Remove the supernatant and resuspend the cell pellet in 100 µL of lysis buffer by gentle pipetting. Incubate for 5 min at room temperature. At this stage, the cell lysates can be stored at –70°C, or alternatively proceed directly to the assays.

### 3.3. β-Galactosidase and Luciferase Assays

Centrifuge the lysed cells from **Subheading 3.2., step 3** at 18,000*g* for 1 min at room temperature to pellet the cell debris. Place the tubes on ice and keep them on ice for the rest of the analysis.

In order to compare different transfections with each other, the enzyme activity present in the same volume of lysate from each transfection has to be compared. However, the β-galactosidase can be measured from one volume and the luciferase from a different volume.

#### 3.3.1. β-Galactosidase Assay

1. β-galactosidase catalyses the hydrolysis of ONPG to o-nitrophenol, which is yellow. Aliquot 250 µL ONPG solution (0.8 mg/mL in β-galactosidase assay buffer) to 1.5-mL tubes, preparing one tube more than the number of transfections.
2. Add 20 µL of the supernatant of each centrifuged lysate to a tube with ONPG solution. Mix well.
3. Incubate at room temperature until the samples are light yellow. Depending on the cell line, this will take anywhere from 5 min to overnight. If no yellow color appears overnight, your transfection has been unsuccessful. If this is the case, repeat **Subheading 3.1.** with different conditions (*see* **Notes 3–6**).
4. To directly compare the amount of β-galactosidase activity in each sample, stop the reactions of all samples after the same incubation time. When yellow color is reached in all samples, stop the reactions by adding 250 µL of 1 *M* $Na_2CO_3$. Also add 250 µL of 1 *M* $Na_2CO_3$ to the extra "blank" tube prepared in **step 1**. Mix well.
5. Measure the optical density at 420 nm ($OD_{420}$) of the reactions vs the "blank." The β-galactosidase-catalyzed reaction is linear and can be accurately measured between an $OD_{420}$ of 0.2 and 0.8. If the $OD_{420}$ of your samples lie outside of this

range, repeat **steps 1–4** and leave the reactions for a longer time or analyze a smaller volume of each lysate (*see* **Note 8**).

### 3.3.2. Luciferase Assay

Luciferase catalyzes the oxidation of its substrate luciferin, which leads to an emission of light. To carry out this reaction, use a commercially available kit (*see* **Subheading 2.3.**). The following protocol is a general outline of the procedure, but refer to the protocol that comes with your kit for specific details.

1. Prepare enough luciferase assay reagent for your samples. Equilibrate this to room temperature, the optimal temperature for luciferase activity measurements.
2. Aliquot 20 μL of the supernatant of each centrifuged lysate to a luminometer cuvet and leave at room temperature for 5 min to equilibrate.
3. Taking one cuvet at a time, add the recommended volume of luciferase assay reagent. Immediately measure the light emission of the reaction by placing the cuvet in a luminometer. It is important to transfer the cuvet to the luminometer as soon as possible after adding the reagent because the light intensity of the reaction is constant for only a few seconds (specified in your kit information) and then begins to decay.

## 3.4. Analysis of Data

The luciferase activity value is a measure of transcription initiated from the reporter plasmid. However, the luciferase value of one transfection cannot without precaution be directly compared with the luciferase value of another, because the cells might not have received an identical amount of DNA during the electroporation. The β-galactosidase value is a measure of the efficiency of an individual transfection. Therefore, to compare different transfections, normalize the luciferase value of each individual transfection by dividing it by its corresponding β-galactosidase value. To determine the effect of (over-) expressing the cDNA, compare the luciferase/β-galactosidase value of that transfection to the luciferase/β-galactosidase value of cells transfected with an empty expression plasmid.

## 4. Notes

1. When choosing the promoters of the expression and internal control plasmids, consider the possibility that the protein you wish to (over-) express may influence transcription from the promoter. It is best to avoid such a promoter.
2. Although many electroporation cuvets are recommended for single use only, they can be reused up to 10 times with little effect on transfection efficiency. After use, cuvets are thoroughly rinsed in water and stored dry. Sterilize them by standing in a beaker of 70% ethanol a few hours before use, and then allow them to air dry (approx 30 min) in the hood prior to adding cells.

3. The optimal amount of expression plasmid depends on the cell type and the efficiency of the transfection and the expression of the plasmid in these cells. It might also depend on effects of the introduced plasmid or the expressed protein, for example if the protein becomes toxic to the cell above a certain concentration. Therefore try a range of DNA concentrations to determine the optimal concentration for your purposes.
4. The cells can alternatively be resuspended in PBS or PBS supplemented with 20 m*M* HEPES, pH 7.1, for the electroporation procedure. HEPES provides extra buffering capacity to reduce the pH change that occurs at the electrodes, a cause of cell death during electroporation. There appears to be little difference in transfection efficiency between PBS, PBS-HEPES, or culture medium, but cells may show enhanced survival when electroporated in culture medium. Transfection efficiency can be affected by the salt concentration of the electroporation buffer ***(12)***. Furthermore, addition of carrier DNA can improve transfection efficiency ***(12)***. It is therefore recommended to use carrier DNA, such as an inert plasmid, when the amount of expression plasmid is low.
5. For cells more resistant to transfection, efficiency may be improved by incubating the cells (in the cuvets) on ice for 5 min before and after electroporation. This may reduce the kinetics of pore closure and thus provide more time for the DNA to enter the cell. It may also protect the cells from heat damage when subjected to the electric pulse.
6. The voltage setting is the most critical parameter of the transfection procedure. Too low a voltage will have no effect on the cell membrane, but too high a voltage irreversibly damages the cell. The optimal voltage should be determined empirically using your experimental conditions. As a guideline, **Table 1** lists voltages that are successful for a variety of cell lines using the indicated electroporation system. Comprehensive studies of optimal voltages for other cell types appear in the literature (for example, **refs. *4,12–14***).
7. Transfected cells can be analyzed several hours to several days after electroporation. Times shorter than 8 h are usually not long enough for reasonable expression of the protein. Over a period of several days, the cells lose transfected DNA and thus expression from the plasmid will gradually decrease. For studies of expression over a longer period of time, stable transfectants are needed. Albeit at a very low frequency, exogenously introduced DNA can insert into the chromosomal DNA. If the transfected expression plasmid also encodes a selectable marker, for example a drug resistance gene, it is possible to select for and amplify these cells ***(15–17)***. These stable transfectants can continue to express the integrated gene for an indefinite time.
8. If the β-galactosidase activity of the samples varies more than fourfold, it is not possible to stop the reactions when all samples lie in the $OD_{420}$ range of 0.2–0.8. In this case, stop individual reactions when they have reached the appropriate color, record the reaction time, and wait until all reactions are finished before measuring the $OD_{420}$ of the samples. The ONPG hydrolysis per time unit is then calculated.

**Table 1**
**Examples of Suitable Voltages for Electroporation Using the Gene Pulser® II Electroporation System (Bio-Rad) with a Capacitance of 950 μF and a Time Constant of 15–20 ms**

| Cell line | Voltage | Cell line | Voltage |
|---|---|---|---|
| MOLT-4 | 290 | Jurkat | 330 |
| DG-75 | 300 | K-562 | 340 |
| HL-60 | 300 | Primary B-cells | 340 |
| Raji | 300 | CTLL-2 | 360 |
| BL-41 | 320 | EL4 | 370 |

## Acknowledgments

This work was supported by grants from the Swedish Cancer Society, the Swedish Research Council for Engineering Sciences/SSF, and the Swedish Natural Science Research Council.

## References

1. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Expression of cloned genes in cultured mammalian cells, in *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 16.1–16.81.
2. Keown, W. A., Campbell, C. R., and Kucherlapati, R. S. (1990) Methods for introducing DNA into mammalian cells. *Methods Enzymol.* **185,** 527–537.
3. Spector, D. L., Goldman, R. D., and Leinwand, L. A., eds. (1998) Preparation of macromolecules and introduction into cells, in *Cells: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 82.1–93.22.
4. Potter, H. (1988) Electroporation in biology: methods, applications, and instrumentation. *Anal. Biochem.* **174,** 361–373.
5. Tsong, T. Y. (1991) Electroporation of cell membranes. *Biophys. J.* **60,** 297–306.
6. Rasmussen, C. D., Means, R. L., Lu, K. P., May, G. S., and Means, A. R. (1990) Characterization and expression of the unique calmodulin gene of Aspergillus nidulans. *J. Biol. Chem.* **265,** 13,767–13,775.
7. Colomer, J., Agell, N., Engel, P., and Bachs, O. (1994) Expression of calmodulin and calmodulin binding proteins in lymphoblastoid cells. *J. Cell Physiol.* **159,** 542–550.
8. Ye, Q., Wei, Y., Fischer, R., Borner, C., and Berchtold, M. W. (1997) Expression of calmodulin and calmodulin binding proteins in rat fibroblasts stably transfected with protein kinase C and oncogenes. *Biochim. Biophys. Acta* **1359,** 89–96.
9. Zheng, Q. and Chang, D. C. (1991) High-efficiency gene transfection by *in situ* electroporation of cultured cells. *Biochim. Biophys. Acta* **1088,** 104–110.
10. Raptis, L. H., Firth, K. L., Brownell, H. L., Todd, A., Simon, W. C., Bennett, B. M., et al. (1995) Electroporation of adherent cells *in situ* for the introduction of nonpermeant molecules. *Methods Mol. Biol.* **48,** 93–113.

11. Bright, G. R., Kuo, N.-T., Chow, D., Burden, S., Dowe, C., and Przybylski, R. J. (1996) Delivery of macromolecules into adherent cells via electroporation for use in fluorescence spectroscopic imaging and metabolic studies. *Cytometry* **24,** 226–233.
12. Chu, G., Hayakawa, H., and Berg, P. (1987) Electroporation for the efficient transfection of mammalian cells with DNA. *Nucleic Acids Res.* **15,** 1311–1326.
13. Knutson, J. C. and Yee, D. (1987) Electroporation: parameters affecting transfer of DNA into mammalian cells. *Anal. Biochem.* **164,** 44–52.
14. Andreason, G. L. and Evans, G. A. (1989) Optimization of electroporation for transfection of mammalian cell lines. *Anal. Biochem.* **180,** 269–275.
15. Kaufman, R. J. (1990) Selection and coamplification of heterologous genes in mammalian cells. *Methods Enzymol.* **185,** 537–566.
16. Kane, S. E. (1997) Selection of transfected cells and coamplification of transfected genes. Methods Mol. Biol. **62,** 359–367.
17. Rose, J. (1998) Transfection of DNA into mammalian cells, in *Cells: A Laboratory Manual* (Spector, D. L., Goldman, R. D., and Leinwand, L. A., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 86.1–86.6.

# 28

# Monitoring the Intracellular Free $Ca^{2+}$-Calmodulin Concentration with Genetically-Encoded Fluorescent Indicator Proteins

**Anthony Persechini**

## 1. Introduction

Calmodulin (CaM) is probably the single most important $Ca^{2+}$-binding protein in the cell by virtue of its central role in converting $Ca^{2+}$ signals into biochemical events. It accomplishes this conversion primarily by controlling in a $Ca^{2+}$-dependent manner the activities of a number of different target proteins. Particularly well-studied examples include the myosin light-chain kinases, CaM kinases I, II, and IV, calcineurin, the constitutive nitric oxide synthases, adenylyl cyclases I and VIII, and the cyclic nucleotide phosphodiesterases ***(1–6)***. In general, CaM is thought to remain dissociated from its targets at resting free-$Ca^{2+}$ concentrations. The protein contains four EF-hand $Ca^{2+}$-binding domains, and it must bind three to four $Ca^{2+}$ ions before activating a typical target protein, such as myosin light-chain kinase or phosphodiesterase ***(7,8)***. There are several exceptions to this overall picture: proteins containing IQ-motifs, such as neuromodulin or unconventional myosins ***(9–11)***, bind $Ca^{2+}$-free CaM as well or better than $Ca^{2+}$-liganded CaM, and CaM is an integral subunit in several proteins, including ryanodine receptors, small conductance potassium channels, inducible nitric oxide synthase, and phosphorylase *b* kinase ***(12–15)***.

In spite of these exceptions, it is clear that for many critically important targets, the free concentration of $Ca^{2+}$–CaM ($[Ca^{2+}–CaM]_i$) is a crucial determinant of activity. The $[Ca^{2+}–CaM]_i$ produced at a particular intracellular free-$Ca^{2+}$ ion concentration is not easily inferred from in vitro data. It is determined by the amounts and distributions of targets and CaM, the affinities

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel © Humana Press Inc., Totowa, NJ

of the different CaM-target complexes, and thermodynamic coupling effects on the $Ca^{2+}$-binding affinity of the CaM-target complexes. Furthermore, the amount and subcellular distribution of CaM appears to change during the cell cycle and in response to some stimuli ***(16–18)***, and different cell types differ greatly in the total amounts of CaM and targets present ***(19,20)***. The distributions of some CaM targets, such as CaM kinase II and calcineurin, also can change dynamically ***(21,22)***. Hence, the $[Ca^{2+}–CaM]_i$ produced at a particular $[Ca^{2+}]_i$ is likely to vary among different cell types, and spatiotemporally within cells of the same type. It is clear that a method to monitor $[Ca^{2+}–CaM]_i$ in intact cells is required before we can investigate transduction of intracellular $Ca^{2+}$ signals by CaM.

We have developed a family of fluorescent indicators for $Ca^{2+}$–CaM whose responses are based on CaM-dependent changes in fluorescence resonance energy transfer (FRET) between variants of green fluorescent protein (GFP) ***(23,24)***. This family of indicators can be stably expressed in mammalian cells, so that $[Ca^{2+}–CaM]_i$ can be monitored in living cells without the preliminary manipulations needed with organic indicators, such as those typically used to monitor $[Ca^{2+}]_i$. As we have reported elsewhere ***(25)***, if CaM is fused to these indicator constructs they become directly responsive to changes in $[Ca^{2+}]_i$. Tsien and co-workers ***(26)*** have used a similar approach to develop GFP-based $Ca^{2+}$ indicators they term "cameleons." Whereas GFP-based $Ca^{2+}$ indicators clearly have significant utility, they are not the subject of this chapter. We will focus here on the methods used to construct, characterize, and express indicators for $Ca^{2+}$–CaM, and to utilize them to monitor $[Ca^{2+}–CaM]_i$ in living cells.

## 2. Materials

Materials are generally given within the protocol descriptions, where they can be understood in the context of the procedures in which they are employed. Additional details concerning the materials used are presented here. Unless otherwise stated, all reagents are obtained from standard sources and are of analytical grade. Double-deionized water is used throughout. Equipment and materials required for routine molecular biology procedures are not listed, as these procedures are not detailed in this chapter.

1. Pure CaM: We use calmodulin expressed in *Escherichia coli* in all our protocols. We have described the procedures for purification of bacterially expressed CaM in detail elsewhere ***(27)***. Purified CaM is also commercially available from a number of sources.
2. Sequences encoding EYFP, ECFP, and CaM-binding linkers: DNA sequences encoding the GFP variants are available commercially from Clontech, Inc. (Palo Alto, CA). DNA sequences encoding the CaM-binding linker sequences were

constructed from overlapping oligonucleotides, and are based on the CaM-binding sequence in avian smooth muscle myosin light-chain kinase ***(28)***.
3. BAPTA: (1,2,-*bis*(o-aminophenoxy)ethane-*N*,*N*,*N*,'*N*'-tetraacetic acid) is obtained from Molecular Probes, Inc. (Eugene, OR).
4. β-escin and α-toxin (α-hemolysin) are both obtained from Sigma (St. Louis, MO); the properties of these reagents can be variable, so the efficacy and optimal concentration should be determined for each batch.
5. Terrific broth: 10.0 g tryptone, 20 g yeast extract, 2.65 g $KH_2PO_4$, 4.33 g $Na_2HPO_4$, and 4 mL glycerol/L.
6. Resuspension buffer: 50 m*M* Tris-HCl, 0.1 *M* NaCl, 1 m*M* EDTA, pH 8.0.
7. Hen egg-white lysozyme.
8. DNase buffer: 15 ng/mL DNase I and 3 m*M* $MgCl_2$.
9. Column buffers:
a. 50 m*M* Tris-HCl, 0.1 *M* NaCl; pH 8.0.
b. 50 m*M* Tris-HCl, 0.5 *M* NaCl, 5 m*M* imidazole; pH 8.0.
c. 50 m*M* Tris-HCl, 0.1 *M* NaCl; pH 8.0.
10. Elution buffer: 50 m*M* Tris-HCl, 0.1 *M* NaCl, 100 m*M* imidazole; pH 8.0.
11. Dialysis buffer: 10 m*M* Tris-HCl, 0.1 *M* KCl; pH 8.0.

## 3. Methods

The CaM indicators in use in our laboratory have undergone significant improvements since our initial report. In particular, we now utilize the enhanced cyan (ECFP) and yellow (EYFP) GFP color variants described by Tsien and coworkers ***(26,29,30)*** as a FRET pair in place of the blue and red color variants, which are not suitable for emission ratio measurements. The fluorophore in the blue variant is highly susceptible to photobleaching, and has a significantly lower quantum efficiency than other GFP fluorophores. We also have altered the linker sequence between the two GFPs to produce indicators with several different affinities for $Ca^{2+}$–CaM. The constructs currently in use are presented schematically in **Fig. 1**. We designate specific CaM indicators as FIP-$CB_x$, where "FIP" stands for "fluorescent indicator protein," "CB" stands for "CaM binding," and "*X*" is an identifying subscript.

### *3.1. Purification and Characterization of Engineered CaM Indicators*

Genes encoding all the CaM indicators are first assembled in a pET30a (Novagen, Inc.) bacterial expression vector, modified to remove a single endogenous *Kpn*I restriction site to facilitate subsequent cloning procedures. A map for one of these constructs, pETIC-35, is presented in **Fig. 2**. This particular vector is for expression of FIP-$CB_{SM-35}$, an ECFP/EYFP-based indicator containing an unmodified CaM linker (L1 in **Fig. 1**). Expression is under control of a *T7/lac* hybrid promoter, which requires a bacterial host harboring a

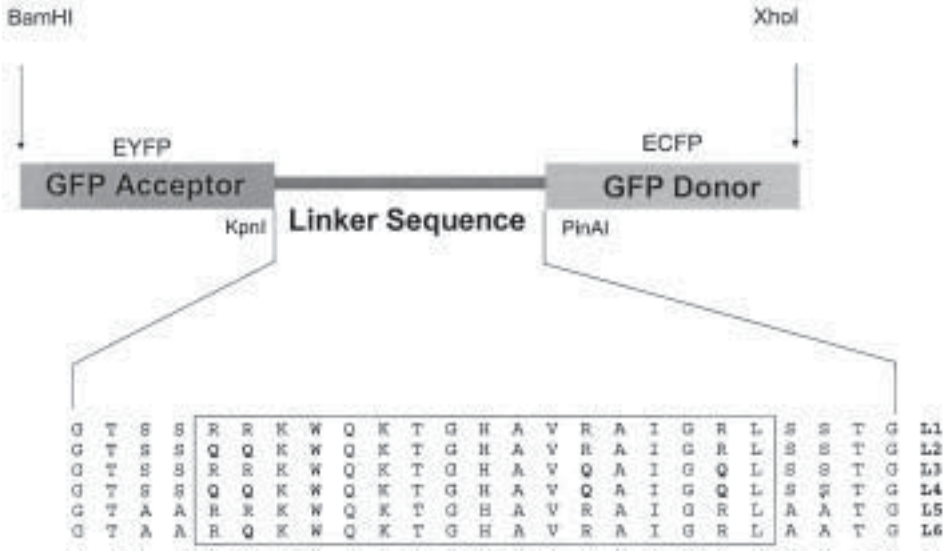


Fig. 1. Schematic representation of indicators for monitoring $[Ca^{2+}–CaM]_i$. The DNA encoding the indicators is constructed in a *Bam*HI–*Xho*I cassette to facilitate transfer between vectors for expression in *E. coli* and in mammalian cells. The $Ca^{2+}$–CaM-binding linker sequence is based on the CaM-binding domain in avian smooth muscle light chain kinase ***(28)***. Several variants have been made by substituting essential arginine residues with glutamine residues. We have also made variants in which the flanking serine residues in the original construct, which play no role in CaM binding, have been replaced by alanines. This was done because these serines (especially those on the C-terminal side of the linker) are potential candidates for protein kinase-catalyzed phosphorylation. We have since observed no differences between the behavior of indicators containing linkers with flanking alanines and those with flanking serines either in vitro or when expressed in mammalian cells. The $K_d$ values determined for indicators constructed with the different linker sequences shown are: L1; <1 n*M*, L2; 45 n*M*, L3; 400 n*M*, L4; approx 10 μ*M*, L5; <1 n*M*, L6; 2 n*M*.

gene for T7 RNA polymerase. We utilize *E. coli* BL21(DE3), which is lysogenic for λDE3, and is also deficient in lon and *ompT* proteases. Expression is induced by adding isopropyl-β-D-thiogalactoside (IPTG), a galactose analog that derepresses expression of both the polymerase and the cloned gene. To facilitate purification the indicator is expressed in bacteria as a fusion protein with an N-terminal 6-*His* tag (**Fig. 2**).

### *3.1.1. Expression of Indicator Proteins in* E. coli

Some of the older GFP spectral variants can be difficult to express in bacteria. The method we describe here was originally designed to give optimal yields

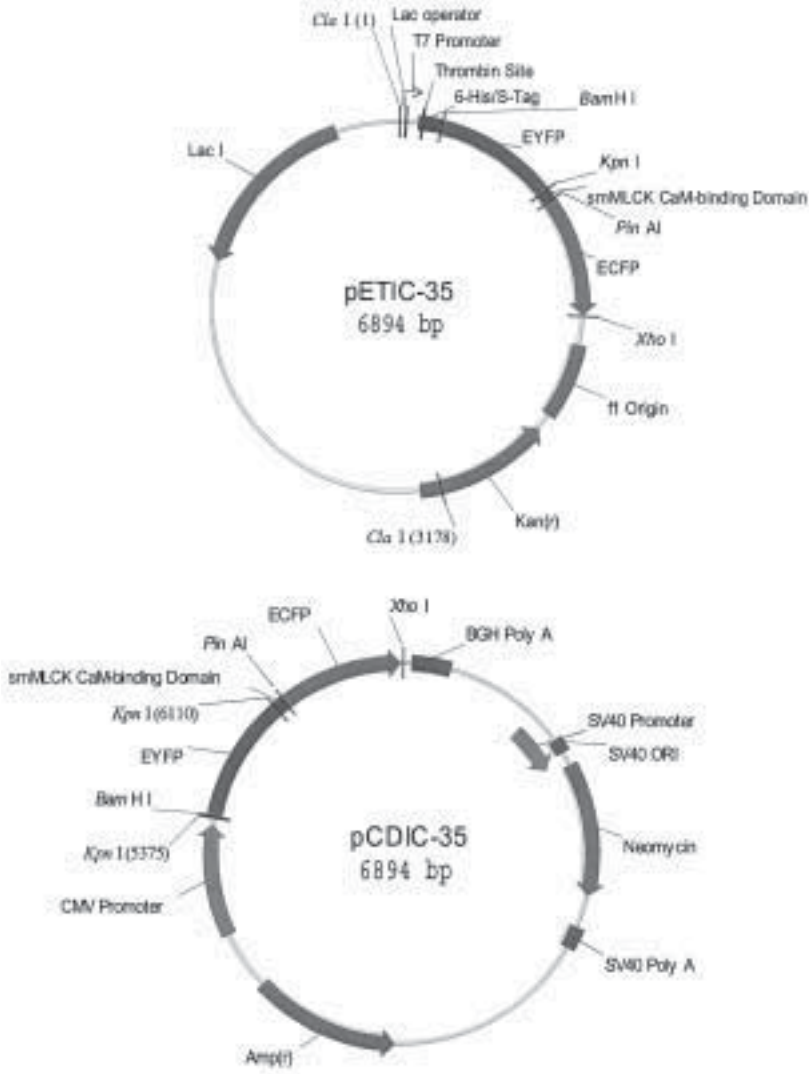


Fig. 2. Maps of vectors for expression of indicators in bacterial (*upper map*) or mammalian (*lower map*) cells. The positions of features and key restriction sites are indicated.

of these difficult variants, but we have found it to give reliable yields of all GFP-based indicators.

1. 50 mL of Terrific broth is inoculated with BL21(DE3) cells transformed with the desired plasmid, and the cells are grown under 50 μg/mL kanamycin selection overnight at 37°C in a shaker running at 300 rpm.

2. 1 L of Terrific broth in a 2800-mL baffled flask (Bellco Glass, Inc., Vineland, NJ) is inoculated with the 50-mL overnight culture. It is not necessary to include kanamycin at this stage. Cells are incubated with shaking at 23°C for 6 h or until the culture has reached an $OD_{660}$ of 0.6–0.8, and expression is induced by adding 0.5 m*M* IPTG. Incubation at 23°C is continued for approx 40 h.
3. The bacterial cells are pelleted at 5000 rpm for 10 min in a Sorvall GS-3 rotor at 4°C, and the cell pellet is resuspended in 100 mL of 50 m*M* Tris-HCl; pH 8.0. The centrifugation step is repeated, and the cell pellet is resuspended in 100 mL of resuspension buffer.
4. 200 μg/mL egg-white lysozyme is added to the cell suspension and it is immediately transferred to centrifuge tubes and incubated on ice for 30 min.
5. The cells are then sonicated for 2 min using a Branson Sonifier 450 with microtip attachment. We use a 40% duty cycle and the maximum microtip power output. The lysed cells are then subject to centrifugation for 20 min at 10,000 rpm in a Sorvall SS-34 rotor at 4°C.
6. DNase buffer is added to the supernatant fraction, followed by centrifugation at 25,000 rpm for 60 min in a Beckman SW27 rotor at 4°C. DNase I is included so that nucleic acids do not clog the affinity column during the final purification step. The supernatant fraction from this procedure is suitable for a preliminary characterization of the indicator fluorescence, and culture volumes can be reduced by a factor of at 20–30 if only the crude supernatant fraction is required.
7. A 1-mL Pharmacia (Uppsala, Sweden) HiTrap® metal chelating column is washed with 5 mL water, primed with 0.5 mL of 0.1 *M* NiCl, and washed with an additional 5 mL water. The supernatant fraction is then pumped onto the column. A greenish-yellow accumulation of bound indicator on the column should develop during this process.
8. The column is washed with 25 mL of column buffer "a," followed by 25 mL column buffer "b," and 15 mL of column buffer "c."
9. The bound indicator is eluted into a minimal volume of elution buffer and dialyzed exhaustively against dialysis buffer.
10. The concentration of indicators constructed with the ECFP/EYFP FRET pair are determined by optical absorbance using a $\varepsilon_{513}$ of 89 m*M*/cm. We no longer routinely perform a gel electrophoretic analysis of the final product, but in the past have consistently observed a purity level ≥ 80%, with the highest purity observed when the affinity column is saturated with 6-*His*-tagged protein prior to elution, which seems to discourage nonspecific binding. After dialysis, indicator proteins can be stored frozen at –80 °C for several months.

### 3.1.2. Spectral Characterization and $K_d$ Determination

For each indicator, it is important at least to determine the changes in the excitation and emission spectra that occur upon binding $Ca^{2+}$–CaM, and the apparent dissociation constant for $Ca^{2+}$–CaM. It is also important to verify the $Ca^{2+}$-dependence of its interaction with CaM. All in vitro fluorescence measurements are performed using a Photon Technologies International

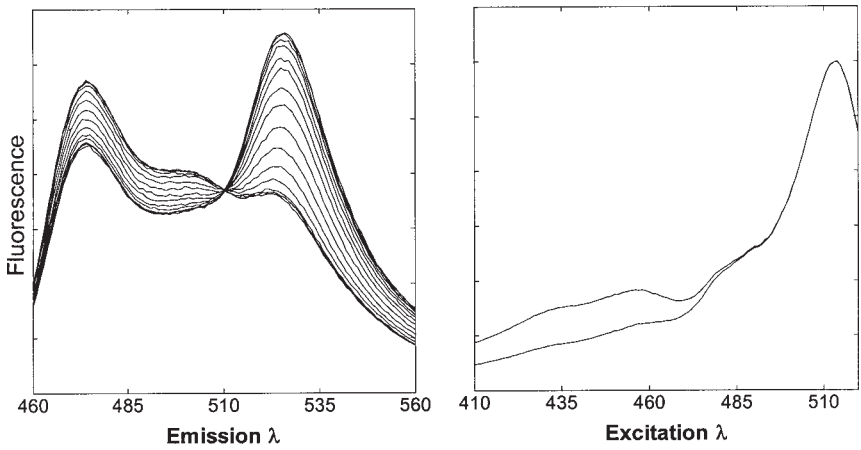


Fig. 3. The effect of $Ca^{2+}$–CaM on the excitation and emission spectra for FIP-$CB_{SM-38}$ ($K_d$ = 45 nm). (*Left panel*) A series of emission spectra obtained after successive additions of pure CaM to a 100 n*M* indicator solution as described in the text. The emission of the EYFP acceptor, centered at approx 535 nm, decreases with each addition until the indicator is saturated. The emission of the ECFP FRET donor, centered at approx 480 nm, behaves in a reciprocal manner. Indicator fluorescence was excited at 430 nm. (*Right panel*) The corrected excitation spectra corresponding to the CaM-free and CaM-saturated indicator are shown. EYFP acceptor emission was monitored at 530 nm.

(Monmouth Junction, NJ) QM–1 photon-counting fluorometer. Excitation and emission spectra are corrected for monochromator artifacts using data supplied by the manufacturer. A relative correction for wavelength-dependent variations in illumination intensity is also applied to excitation spectra. Samples are incubated in a 1 cm × 1 cm fused silica cuvet held at 25°C in a water-jacketed cuvet holder equipped with a magnetic stirrer. In general, 5-nm slits are used on both the emission and excitation monochromator input and output light paths. Spectra for titration of pure FIP-$CB_{SM-35}$ with $Ca^{2+}$–CaM are shown in **Fig. 3**, and binding isotherms for indicators with three different $K_d$ values for $Ca^{2+}$–CaM are presented in **Fig. 4** (*see* **Note 1**).

1. After determination of background fluorescence and water Raman scattering, indicator is added to a cuvet containing 2 mL of 25 m*M* Tris-HCl, 0.1 *M* KCl, 0.5 m*M* $MgCl_2$; pH 7.4 to produce the desired final concentration. If fluorescence data are to be used for a $K_d$ determination, then the lowest acceptable indicator concentration should be used.
2. Small aliquots (approx 2 μL) of a concentrated CaM solution (approx 10 μ*M*) are added directly to the cuvet. Spectra are taken after each addition. It is not necessary to correct for the small changes in total volume. Although contaminating

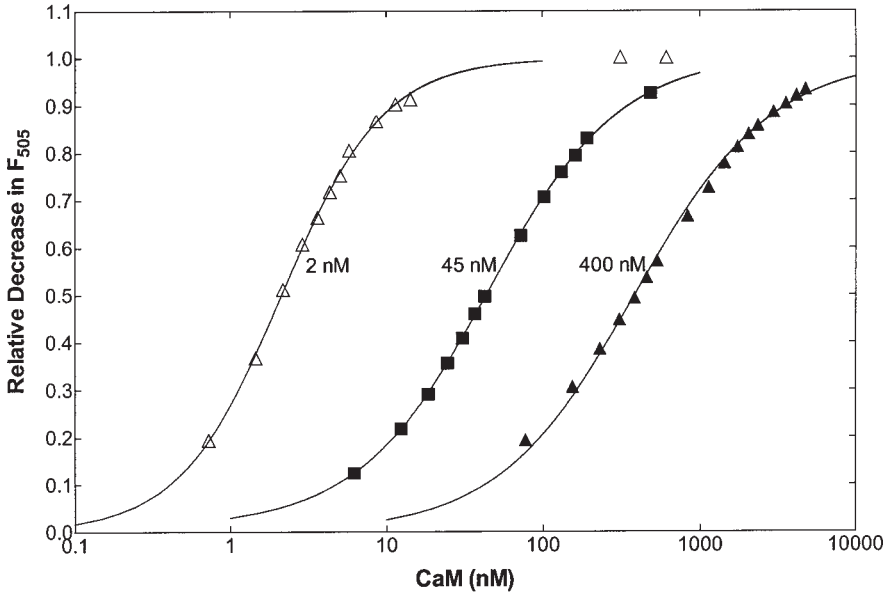


Fig. 4. Titration with CaM of the acceptor fluorescence of indicators constructed using L6 (▲), L2 (■) or L3 (▲) linkers. These data were determined using indicators based on the EBFP/EGFP FRET pair that is no longer used in the laboratory. Hence, the acceptor fluorescence emission at 505 nm was monitored, and fluorescence was excited at 380 nm. We have found that the GFP variants used to construct an indicator have no effect on its affinity for $Ca^{2+}$–CaM; this is appears to be determined solely by the linker. Data were fit to **Eq. 1** (■,▲) or **Eq. 2** (▲), and the derived apparent $K_d$ values are given in the figure.

levels of $Ca^{2+}$ (approx 5 μ*M*) should be sufficient to saturate the $Ca^{2+}$-binding sites on CaM when it is bound to indicator, we add 0.1 m*M* $CaCl_2$ to ensure that the free-$Ca^{2+}$ concentration is not limiting. CaM is adsorbed to surfaces, especially plastics, and this can result in significant losses from dilute solutions. We add 0.1 mg/mL BSA to all buffers to prevent this.

3. 5–10 m*M* BAPTA may be added at the end of an experiment to verify that the interaction between CaM and the indicator is wholly $Ca^{2+}$-dependent. A high CaM concentration can also be added to investigate the possibility of a low-affinity $Ca^{2+}$-independent interaction with the indicator. We have seen no evidence of such an interaction with any CaM indicator at CaM concentrations as high as 10 μ*M*, which is similar to estimates of the *total* CaM concentration in fibroblasts ***(19)***.
4. If the bound and free concentrations of $Ca^{2+}$–CaM are approximately the same ("Michaelis-Menten" conditions), then data for the fractional change in fluorescence produced by each addition of $Ca^{2+}$–CaM can be fit to **Eq. 1**.

$$\frac{F_{max} - F}{F_{max} - F_{min}} = \frac{[CaM]_{tot}}{[CaM]_{tot} - K_d} \quad (1)$$

If Michaelis-Menten conditions do not apply, then the data must be fit to **Eq. 2**.

$$\frac{F_{max} - F}{F_{max} - F_{min}} = \frac{[I]_{tot} - [CaM]_{tot} - K_d - \sqrt{([I]_{tot} + [CaM]_{tot} + K_d)^2 - 4[I]_{tot}[CaM]_{tot}}}{1[I]_{tot}} \quad (2)$$

For example, FIP-$CB_{SM-35}$ has a $K_d$ of 1 n*M* for $Ca^{2+}$–CaM, but in our hands, 2 n*M* is the minimum indicator concentration producing acceptable fluorescence data. Thus, over most of the range of added CaM concentrations the bound and free concentrations of the protein are quite different, so a quadratic is required to fit the data (*see* **Eq. 2**). A drawback to using **Eq. 2** is that it requires a precise knowledge of the indicator concentration, so it is important to perform titrations at two or three different indicator concentrations to ensure that consistent results are obtained. In addition, we should emphasize that little useful information about binding can be extracted from an essentially linear isotherm, such as would be obtained if we were to titrate the response of FIP-$CB_{SM-35}$ at a concentration of 100 n*M*. Binding data are fit directly to **Eq. 1** or **2** using a standard nonlinear least-squares analysis. $F_{max}$ and $F_{min}$ are fluorescence emission measurements made at the acceptor (EYFP) emission maximum (approx 530 nm) when the indicator is CaM-free and CaM-saturated, respectively. $[I]_{tot}$ and $[CaM]_{tot}$ are the total concentrations of indicator and CaM (*see* **Table 1**).

### 3.2. Stable Expression of Indicators in Mammalian Cells

The procedures described here have been developed using HEK-293 cells, a line derived from human embryonic kidney epithelium (ATCC #1573). To construct mammalian expression vectors indicators, we simply excise the DNA encoding it from the bacterial expression vector using *Bam*HI and *Xho*I and ligate the fragment into a pcDNA3 vector (Invitrogen, Inc., Carlsbad, CA) cleaved with these enzymes. Expression of the cloned indicator in these vectors is under control of a cytomegalovirus (CMV) promoter, and a nonfusion protein is produced. The supplier provides variants of this vector that carry selectable markers for zoecin, G418 or blasticidin. A vector map for the construct used to express FIP-$CB_{SM-35}$ is shown in **Fig. 2**.

#### 3.2.1. Transfection and Selection to Produce HEK–293 Cells Stably-Expressing CaM Indicators

LipofectAMINE® (Life Technologies, Inc., Gaithersburg, MD) is used to introduce vector DNA into HEK-293 cells essentially as described by the manufacturer.

1. Cells are plated at a density of $5 \times 10^5$ per 60-mm dish 2 d before transfection. Cells should be 50–80% confluent at the time of transfection.

**Table 1**
**Parameters for Three Different Indicators Determined as Described in Subheading 3.1.2. ($K_d$ Values) and 3.3.3. ($R_{min}$ and $R_{max}$ Values)**

| Indicator | $K_d$ (n*M*) | $R_{min}$ | $R_{max}$ |
|---|---|---|---|
| FIP-$CB_{SM-41}$ | 2 | 0.75 | 1.52 |
| FIP-$CB_{SM-38}$ | 45 | 0.77 | 1.53 |
| FIP-$CB_{SM-39}$ | 400 | 0.77 | 1.49 |
| FIP-$CA_{37}$ | — | 0.89 | 1.5 |

2. For each dish of cells, prepare Solution A (5 μg plasmid DNA in300 μL serum-free media) and Solution B (20 μL lipofectAMINE® in 300 μL serum-free culture media).
3. Solutions A and B are combined with gentle mixing (do not vortex) and incubated at room temperature for 20 min.
4. 2.4 mL of serum-free media is added to the lipid mixture and it is immediately placed on cells that have been rinsed with serum-free media. The cells are then incubated for approx 12 h at 37°C in a humidified 5% $CO_2$ incubator.
5. The DNA/lipid mixture is replaced with 6 mL of normal growth media containing 5% fetal bovine serum (Life Technologies 26140-079) and the cells are incubated for 2–3 d.
6. The transfected cells are then lightly trypsinized and transferred to a 75-$cm^2$ cell-culture flask. Drug selection is commenced the following day. Cells transfected with plasmids conferring neomycin resistance are selected using G418 at a concentration of 800 μg/mL. Cells transfected with plasmids conferring blasticidin resistant are selected using this drug at a concentration of 7 μg/mL. (Fresh media and blasticidin should be placed on the cells every 2–3 d as this drug is unstable in the culture medium.)
7. Stably transfected cells are selected within 10–12 d, and can be visualized as discrete areas of growth on the flask. Stable transfectants are propagated as a mixed-clonal population.
8. The success of transfection and selection procedures can be monitored using fluorescence microscopy. A standard FITC filter set is adequate for this purpose (D480/30 exciter, 505DCLP microscope dichroic, D535/40 emitter). We generally find that 50–70% of stably transfected cells express detectable levels of indicator; the rest presumably lack a functional indicator gene.

### *3.2.2. Quantitation of Indicator Expression in Cells*

Confocal fluorescence microscopy is the most convenient method for assessing indicator expression levels in individual cells. We currently use a Noran OZ CLS*M* instrument, with 488-nm excitation light provided by a Argon/Krypton laser. At this excitation wavelength indicator fluorescence emission is

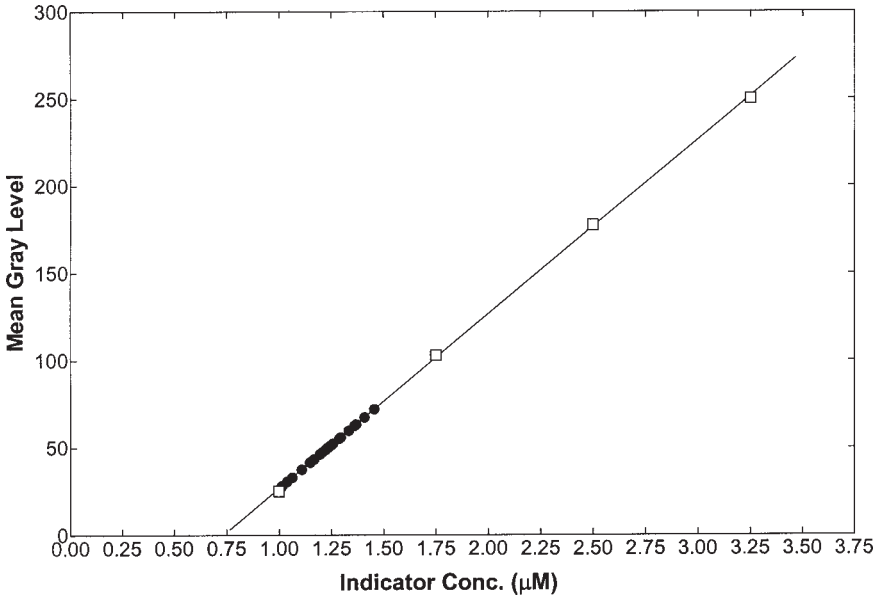


Fig. 5. Estimation of the concentration of expressed indicator in HEK-293 cells using confocal microscopy as described in the text. Mean gray level values for optical sections of pure indicator solutions (□) were plotted vs the indicator concentration to establish the standard curve shown. Estimated indicator concentrations for several cells based on the mean gray level values are also plotted on the standard curve (●).

insensitive to bound CaM (*see* **Fig. 3**). Emitted light passed by a D525/40 filter is detected using a photomultiplier tube.

1. Cells are prepared by growing them on #1 glass cover slips to the desired density. Cover slips are then mounted in a modified Sykes-Moore chamber (Bellco Glass, Vineland, NJ) and overlaid with 1 mL of a standard HEPES buffered saline solution (HBS: 141 m*M* NaCl, 5 m*M* KCl, 1 m*M* $MgSO_4$, 10 m*M* glucose, 10 m*M* HEPES; pH 7.4). Optical sections of 8–12 cells are taken using a 15-μm scanning slit, which is optimal for the ×40 oil-immersion objective used.
2. Using identical settings for gain, offset, and laser power level, optical sections are also taken in standard indicator solutions that have been sandwiched between two #1 glass cover slips mounted in a Sykes-Moore chamber.
3. Emission intensity information digitized as 8-bit gray-level images is analyzed using standard image processing software, and the mean gray-level values for 8–12 regions of interest are determined. We currently use a freely distributed version of NIH Image available from Scion Corporation for this purpose (Frederick, MD). Background subtractions are made using a mean gray-level value calculated based on empty regions of the cover slip.

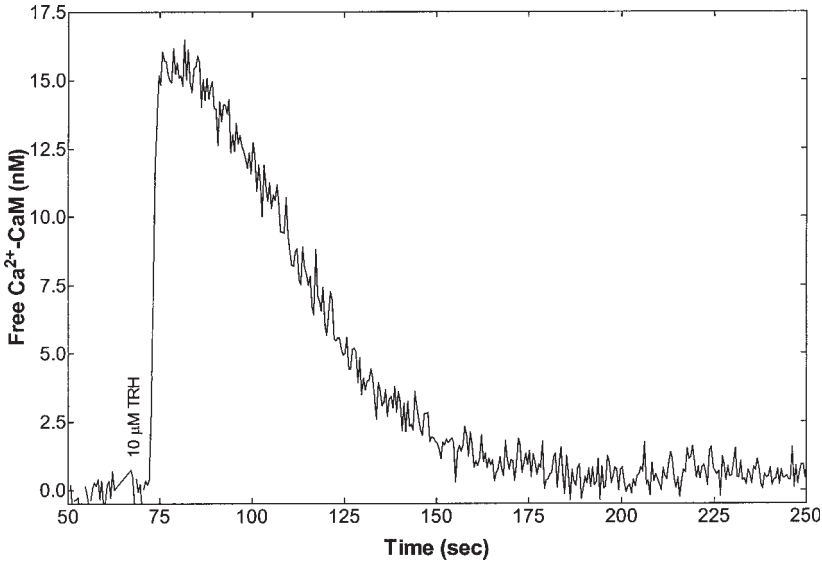


Fig. 6. The $[Ca^{2+}–CaM]_i$ produced in a HEK-293 cell stably expressing FIP-$CB_{SM–38}$ ($K_d$ = 45 n*M*). A transient in $[Ca^{2+}–CaM]_i$ was evoked by adding thyrotropin releasing hormone (TRH) to the cell, which is derived from an HEK-293 cell line (kindly provided by *M*. Shupnick at the University of Virginia) stably expressing the $G_{q/11}$-coupled receptor for this hormone.

A typical set of data obtained using indicator standards and transfected cells is presented in **Fig. 5**. The intercept of the standard curve is nonzero as a result of the PMT offset used. The particular batch of cells represented in the figure expressed indicator at concentrations of 1–2 μ*M*. After determining the expression levels in a population of stably transfected cells, those containing indicator in the desired concentration range can often be selected by eye. The expressed indicators are passively transported into the nucleus, and for unknown reasons are frequently observed to be about twice as concentrated there as in the cytoplasm.

### 3.3. Determining Values for $(Ca^{2+}–CaM)_i$

In this subheading, we describe the procedures used to monitor indicator emission ratios in cells and to calibrate the indicator responses so that values for $[Ca^{2+}–CaM]_i$ can be calculated from them. We employ a microscope photometry system to determine indicator emission ratios in cells, which has the advantages of simplicity, low cost, high sensitivity, and speed, but obviously lacks the spatial resolution available with a slower and much more costly cam-

era-based detection system. A typical $[Ca^{2+}–CaM]_i$ time-course illustrating the effect of a $Ca^{2+}$-mobilizing agonist is presented in **Fig. 6** (*see* **Note 2**).

### 3.3.1. Detection System

The photometry system in our laboratory consists of a Nikon TE200 microscope with a 40× SuperFluor oil immersion objective and dual Hamamatsu R1527P photomultiplier tubes in Model 814 housings mounted on a model D-104C dual-channel photometer (Photon Technologies International, Inc.). Excitation light at 430 nm is supplied by a fiber-optic coupled monochromator with a 75 W xenon arc light source, and is directed through the microscope objective by a 455DCLP dichroic cube. Sample fluorescence passed by the microscope dichroic is split between the two detector channels by a 510DCLP dichroic cube fitted with D535/25 (passed light) and D480/30 (reflected light) filters. Filters and dichroics are obtained from Chroma Technologies (Brattleboro, VT). Fluorescence emission from individual cells is isolated using an adjustable diaphragm at the entrance to the photometer. The digitized output from the PMTs is collected, analyzed, and displayed using the FeliX® software supplied with the photometer.

### 3.3.2. Monitoring Indicator Emission Ratios

It is preferable to monitor emission ratios rather than donor or acceptor emission intensities because ratios are internally normalized for cell-to-cell variations in the amount of indicator expression, and for small changes in the position of a cell image relative to the photometer diaphragm opening that can occur during the course of an experiment.

1. Cells are prepared by plating them in complete medium on sterile #1 glass cover slips at least 1 d before experiments are to be conducted.
2. Cover slips containing cells at the desired density are rinsed in HBS and equilibrated in this buffer for 30 min. We normally include 1 m*M* $CaCl_2$, but it may be omitted, and the equilibration time extended, if $Ca^{2+}$-depleted cells are required.
3. Cover slips are mounted in a Sykes-Moore chamber (Bellco Glass, Inc., Vineland, NJ), overlaid with 1 mL of HBS, and placed in a holder on the microscope stage after preparing the microscope objective with a drop of low-fluorescence immersion oil.
4. After adjusting the photometer diaphragm opening as desired, background values for the two detector channels are measured in an empty region of the cover slip. HEK-293 cells do not exhibit significant autofluorescence at the excitation and emission wavelengths used. These values are automatically subtracted from all subsequent measurements. We usually adjust the photometer diaphragm opening until it is slightly larger than the image of a typical cell, and leave it at this setting throughout an experimental session. New background values must be

determined if the diaphragm opening is changed. It is important to be sure that the PMT output does not exceed its rated maximum, which is approx $2 \times 10^6$ cps for our system.

5. The time-courses for emission intensities at 480 and 535 nm and for the 480/535 emission ratio are continuously monitored. Integration intervals of 0.1 to 0.5 s are used to calculate the PMT output, which limits our time resolution to within this range. For investigations of $[Ca^{2+}–CaM]_i$ in HEK–293 cells, we study the responses to agonists and other agents by adding them directly to the incubation buffer.
6. In situations where emission ratios are static, e.g., under resting conditions or with buffered $[Ca^{2+}]_i$, data from a number of different cells on the same cover slip can be collected simply by moving various individuals into the observation window. We use reticle cross-hairs aligned with the diaphragm opening to facilitate this process.

### 3.3.3. Calibration of Indicator Emission Ratios

To calculate $[Ca^{2+}–CaM]_i$ values based on 480/535 emission ratios we must determine the ratios produced by ligand-free ($R_{min}$) and ligand-saturated ($R_{max}$) indicator (*see* **Notes 2** and **3**).

#### 3.3.3.1. Determination of $R_{min}$ Values

To ensure that a minimal $[Ca^{2+}]_i$ is produced we use a technique involving cell permeabilization with α-toxin, which forms pores in the cell membrane that are permeated only by molecules smaller than approx 3000 Daltons. We have obtained similar $R_{min}$ values using 5 μ*M* ionomycin instead of α-toxin, but prefer to use the toxin because it facilitates verification of $R_{min}$ values, as described below.

1. Cells on a cover slip mounted in a Sykes-Moore chamber are incubated in HBS containing 1 m*M* $CaCl_2$ and 15–30 μg/mL α-toxin on the microscope stage for approx 20 min until the 480/535 emission ratios begin to rise because of entry of $Ca^{2+}$ through the toxin pores.
2. The incubation buffer is then replaced with fresh HBS containing 3 m*M* BAPTA and no added $CaCl_2$, to produce a nominally $Ca^{2+}$-free solution.
3. After the emission ratios have stabilized, they are measured in a number of cells and a mean value for $R_{min}$ is calculated (*see* **Table 1**).
4. To verify that $[Ca^{2+}]_i$ values below those needed to produce a detectable $[Ca^{2+}–CaM]_i$ are present when the $R_{min}$ value is determined, $CaCl_2$ can be incrementally added to the BAPTA/HBS incubation buffer until the 480/535 emission ratios are seen to increase. Allow approx 20 min for equilibration between each addition. If desired, the $[Ca^{2+}]_i$ values after each addition can be estimated using a computer program such as MaxChelator ***(31)***. We have found that there is no significant increase in $[Ca^{2+}–CaM]_i$ until $[Ca^{2+}]_i$ reaches ~ 0.2 μ*M* ***(24)***.

#### 3.3.3.2. Determination of $R_{max}$ Values

To determine $R_{max}$ values cells must be permeabilized with an agent that will allow entry of CaM added to the incubation buffer, while minimizing leakage of the indicator. We have successfully used β-escin for this purpose ***(17,32)***. The in vitro $R_{max}$ values for FIP-CB and FIP-CB-CaM fusions are identical, so we have been able to verify this procedure by comparing the $R_{max}$ values determined for FIP-CBs with the $R_{max}$ value for a FIP-CB-CaM fusion, which can be determined without β-escin permeabilization ***(24)***. Having determined the $R_{max}/R_{min}$ ratio for a mixed-clonal cell line it is not necessary to measure an $R_{max}$ value for each experiment. It can simply be estimated based on the $R_{min}$ value, which, in practice, is usually equivalent to the indicator emission ratio in resting cells.

1. Cells are incubated in HBS containing 3 m*M* BAPTA and 25–50 μ*M* β-escin added from a 5 m*M* stock in ethanol. Successful permeabilization is usually indicated by a transient decrease in fluorescence at 535 nm, which is presumably due to a pH transient.
2. After cells are permeabilized, 10 μ*M* CaM and 5 m*M* $CaCl_2$ are added and 480/535 emission ratios are monitored until they reach a maximum value, which in our hands is consistently twofold larger than the $R_{min}$ value. Emission ratios from a number of cells are averaged to determine the $R_{max}$ value. We have observed considerable cell-to-cell variability with respect to both the incubation time required for permeabilization and the amount of indicator leakage that occurs.

### 3.3.4. Calculation of $[Ca^{2+}–CaM]_i$ Values

The approach used to calculate $[Ca^{2+}–CaM]_i$ values is essentially identical to the approach used with organic "ratiometric" $Ca^{2+}$ indicators ***(33)***. The relationship between $[Ca^{2+}–CaM]_i$ and any given indicator emission ratio is given by

$$[Ca^{2+}–CaM]_i = K'_d\,[R - R_{min} / R_{max} - R] \quad (3)$$

where $K'_d$ is related to the indicator $K_d$ value according to the relation

$$K'_d = K_d\,(S_{f,2} / S_{b,2}) \quad (3)$$

where $S_{f,2}$ and $S_{f,3}$ are the indicator emission intensities at the second wavelength (535 nm for the ECFP/EYFP FRET pair) when the indicator is free or bound. The $S_{f,2}/S_{f,3}$ ratio can be determined using emission intensity data collected when the $R_{max}$ value is measured, we routinely obtain a value of 1.8 (*see* **Notes 4–6**).

## 4. Notes

1. Because we cannot determine indicator $K_d$ values in cells, we must use in vitro $K_d$ values to calculate $[Ca^{2+}–CaM]_i$. The apparent $K_d$ values for the indicators are

relatively insensitive to variations in ionic strength and pH in the physiological range (A. Persechini, unpublished observations). In addition, it is reasonable to assume that $[Ca^{2+}–CaM]_i$ values derived using in vitro indicator $K_d$ values can be used to predict the behavior of CaM targets, whose $K_d$ values also must be determined in vitro.

2. It is important to bear in mind that the uncertainty in the calculated values for $[Ca^{2+}–CaM]_i$ increases rapidly as one approaches either $R_{min}$ or $R_{max}$, because errors in these values and in the measured emission ratio become an increasingly large fraction of the ratio difference (*see* **Eq. 3**).
3. It is difficult to unambiguously determine which $Ca^{2+}$-liganded forms of CaM are reported by indicators expressed in cells. Based on the Hill coefficients for the in vitro dependencies of the indicator responses on the free-$Ca^{2+}$ concentration, $(Ca^{2+})_4–CaM$ appears to be the species reported, in agreement with data for the $Ca^{2+}$-dependence of myosin light-chain kinase activation ***(23,34)***. However, measurements made in cells suggest that it may be $(Ca^{2+})_3–CaM$, or perhaps both $(Ca^{2+})_3–CaM$ and $(Ca^{2+})_4–CaM$ that are reported in this environment ***(24)***. The cell contains a complex multicomponent system of CaM-binding proteins and is spatially heterogeneous in many respects, so interpretation of a Hill coefficient is not straightforward. Thus, although it seems unlikely that the $Ca^{2+}–CaM$ species bound by the indicator changes when it is expressed in a cell, we prefer to use the generic "$Ca^{2+}–CaM$" designation to avoid specifying the particular $Ca^{2+}$-liganded form of CaM bound by the indicator in cells.
4. The EYFP variant used to construct the indicators described here undergoes a pH-dependent reduction in fluorescence emission at 535 nm with an apparent $pK_a$ of approx 7 ***(30)***. The $pK_a$ for this transition can be shifted to a lower value by mutating the EYFP amino sequence so as to stabilize the phenolate form of the chromophore ***(29,30)***. We have produced such a modified EYFP, and find that cytoplasmic indicators containing it are expressed at reduced levels, but otherwise function normally. However, in plasma membrane-targeted constructs the chromophore in the altered EYFP seems to form only when cells expressing it are incubated either overnight at 30°C or for approx 1 h at room temperature. Tsien and co-workers ***(30)*** have encountered similar problems expressing an ER-targeted construct containing the altered EYFP.
5. Introducing an indicator for any ligand into the cell suffers from a generic problem, which is that the indicator itself invariably perturbs the measured values for the free-ligand concentration. Although this problem can never be completely eliminated, its magnitude can be reduced by using an indicator with the lowest affinity possible at the lowest practical concentration. We have found that the lowest practical indicator concentration readily monitored using our photometer-based detection system is approx 1 μ*M*. We have compared the $[Ca^{2+}–CaM]_i$ produced at a saturating $Ca^{2+}$ concentration in cells expressing $Ca^{2+}–CaM$ indicators with $K_d$ values of 2, 45 and 400 nM, and find that the 45 and 400 n*M* $K_d$ indicators report similar values, suggesting that neither greatly perturbs the CaM system. In contrast, the maximum $[Ca^{2+}–CaM]_i$ value reported by the 2 n*M* $K_d$ indicator are

approx 20-fold less. The 45 n*M* $K_d$ indicator appears to represent a good balance between maximizing the indicator response to a typical $[Ca^{2+}–CaM]_i$ transient and minimizing perturbation of the CaM system.

## References


1. Nairn, A. C. and Picciotto, M. R. (1994) Calcium/calmodulin-dependent protein kinases. *Semin. Cancer Biol.* **5,** 295–303.
2. Gallagher, P. J., Herring, B. P., and Stull, J. T. (1997) Myosin light chain kinases. *J. Muscle Res. Cell Motil.*, **18,** 1–16.
3. Bredt, D. S. and Snyder, S. H. (1990) Isolation of nitric oxide synthetase, a calmodulin-requiring enzyme. *Proc. Natl. Acad. Sci. USA* **87,** 682–685.
4. Stemmer, P. M. and Klee, C. B. (1994) Dual calcium ion regulation of calcineurin by calmodulin and calcineurin B. *Biochemistry* **33,** 6859–6866.
5. Cooper, D. M., Mons, N., and Karpen, J. W. (1995) Adenylyl cyclases and the interaction between calcium and cAMP signalling. *Nature* **374,** 421–424.
6. Cox, J. A., Malnoe, A., and Stein, E. A. (1981) Regulation of brain cyclic nucleotide phosphodiesterase by calmodulin. *J. Biol. Chem.* **256,** 3218–3222.
7. Manalan, A. S. and Klee, C. B. (1984) Calmodulin. *Adv. Cycl. Nuc. Prot. Phos. Res.* **18,** 227–278.
8. Vogel, H. J. and Zhang, M. J. (1995) Protein engineering and NMR studies of Calmodulin. *Mol. Cell Biochem.* **149,** 3–15.
9. Apel, E. D. and Storm, D. R. (1992) Functional domains of neuromodulin (GAP-43). *Perspect. Dev. Neurobiol.* **1,** 3–11.
10. Coluccio, L. M. (1997) Myosin I. *Am. J. Physiol.* **273,** C347–359.
11. Whittaker, M. and Milligan, R. A. (1997) Conformational changes due to calcium-induced calmodulin dissociation in brush border myosin I-decorated F-actin revealed by cryoelectron microscopy and image analysis. *J. Mol. Biol.* **269,** 548–557.
12. Cho, H. J., Xie, Q. W., Calaycay, J., Mumford, R. A., Swiderek, K. M., Lee, T. D., and Nathan, C. (1992) Calmodulin is a subunit of nitric oxide synthase from macrophages. *J. Exp. Med.* **176,** 599–604.
13. Shenolikar, S., Cohen, P. T., Cohen, P., Nairn, A. C., and Perry, S. V. (1979) The role of calmodulin in the structure and regulation of phosphorylase kinase from rabbit skeletal muscle. *Eur. J. Biochem.* **100,** 329–337.
14. Xia, X. M., Fakler, B., Rivard, A., Wayman, G., Johnson-Pais, T., Keen, J. E., et al. (1998) Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* **395,** 503–507.
15. Shoshan-Barmatz, V. and Ashley, R. H. (1998) The structure, function, and cellular regulation of ryanodine-sensitive $Ca^{2+}$ release channels. *Int. Rev. Cytol.* **183,** 185–270.
16. Deisseroth, K., Heist, E. K., and Tsien, R. W. (1998) Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons. *Nature* **392,** 198–202.
17. Luby-Phelps, K., Hori, M., Phelps, J. M., and Won, D. (1995) $Ca^{(2+)}$-regulated dynamic compartmentalization of calmodulin in living smooth muscle cells. *J. Biol. Chem.* **270,** 21,532–21,538.

18. Chafouleas, J. G., Bolton, W. E., Hidaka, H., Boyd, A. E. D., and Means, A. R. (1982) Calmodulin and the cell cycle: involvement in regulation of cell-cycle progression. *Cell* **28,** 41–50.
19. Kakiuchi, S., Yasuda, S., Yamazaki, R., Teshima, Y., Kanda, K., Kakiuchi, R., and Sobue K. (1982) Quantitative determinations of calmodulin in the supernatant and particulate fractions of mammalian tissues. *J. Biochem.* **92,** 1041–1048.
20. Vanaman, T. C. and Klee, C. B. (1982) Calmodulin. *Adv. Protein Chem.* **35,** 213–321.
21. Schulman, H., Heist, K., and Srinivasan, M. (1995) Decoding $Ca^{2+}$ signals to the nucleus by multifunctional CaM kinase. *Prog. Brain Res.* **105,** 95–104.
22. Beals, C. R., Clipstone, N. A., Ho, S. N., and Crabtree, G. R. (1997) Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. *Genes Devel.* **11,** 824–834.
23. Romoser, V. A., Hinkle, P. M., and Persechini, A. (1997) Detection in living cells of $Ca^{2+}$-dependent changes in the fluorescence of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. *J. Biol. Chem.* **272,** 13,270–13,274.
24. Persechini, A. and Cronk, B. (1999) The relationship between the free concentrations of $Ca^{2+}$ and $Ca^{2+}$-calmodulin in intact cells. *J. Biol. Chem.* **274,** 6827–6830.
25. Persechini, A., Lynch, J. A., and Romoser, V. A. (1997) Novel fluorescent indicator proteins for monitoring free intracellular $Ca^{2+}$. *Cell Calcium* **22,** 209–216.
26. Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., et al. (1997) Fluorescent indicators for $Ca^{2+}$ based on green fluorescent proteins and calmodulin. *Nature* **388,** 882–887.
27. Persechini, A., Blumenthal, D. K., Jarrett, H. W., Klee, C. B., Hardy, D. O., and Kretsinger, R. H. (1989) The effects of deletions in the central helix of calmodulin on enzyme activation and peptide binding. *J. Biol. Chem.* **264,** 8052–8058.
28. Lukas, T. J., Burgess, W. H., Prendergast, F. G., Lau, W., and Watterson, D. M. (1986) Calmodulin binding domains: characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. *Biochemistry* **25,** 1458–1464.
29. Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem.* **67,** 509–544.
30. Miyawaki, A., Griesbeck, O., Heim, R., and Tsien, R. Y. (1999) Dynamic and quantitative $Ca^{2+}$ measurements using improved cameleons. *Proc. Natl. Acad. Sci. USA* **96,** 2135–2140.
31. Bers, D., Patton, C., and Nuccitelli, R. (1994) A practical guide to the preparation of $Ca^{2+}$ buffers. *Methods Cell Biol.* **40,** 3–29.
32. Brozovich, F. V. (1995) PKC regulates agonist-induced force enhancement in single alpha-toxin permeabilized smooth muscle cells. *Am. J. Physiol.* **268,** C1202–C1206.
33. Grynkiewicz, G., Poenie, M., and Tsien, R. (1985) A new generation of $Ca^{2+}$ indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260,** 3440–3450.
34. Blumenthal, D. K. and Stull, J. T. (1980) Activation of skeletal muscle myosin light chain kinase by calcium$^{(2+)}$ and calmodulin. *Biochemistry* **19,** 5608–5614.

# 29

# Studying the Spatial Distribution of $Ca^{2+}$-Binding Proteins

*How Does it Work for Calmodulin?*

**Katalin Török, Richard Thorogate, and Steven Howell**

## 1. Introduction

Calmodulin is a ubiquitous $Ca^{2+}$-switch protein whose in vitro properties have been widely studied ***(1)***. Visualization of calmodulin levels and functional changes in living cells allows investigations of how calmodulin is involved in organizing specific cellular responses to various stimuli. The advancement of several protein chemistry, biochemical, and microscopic techniques has made the direct study of calmodulin in cellular function less perturbing, more sensitive, and of higher temporal and spatial resolution. For example, more selective fluorescent-labeling techniques directed at strategically positioned Lys and Cys residues in the protein are now available, the latter are introduced by site-directed mutagenesis. In addition, the conjugation of calmodulin c-DNA with enhanced green fluorescent protein (GFP) provides increased sensitivity. Taken together, these advances allow the fluorescence signal of the protein to act as an intracellular reporter group of concentration changes in cell compartments, as well as other well-defined molecular events (e.g., $Ca^{2+}$ and target binding, conformational change). Brighter fluorophores provide increased sensitivity and thus the fluorescent protein may be applied at a lower concentration to act as a tracer of endogenous calmodulin. Calmodulins with probes attached at a single site that have been characterized functionally by comparison to unmodified calmodulin, facilitate clearer data interpretation. Laser-scanning confocal microscopy offers higher resolution so events in living cells can be monitored in greater detail in order to understand calmodulin and its inter-

From: *Methods in Molecular Biology*, vol. 173:
*Calcium-Binding Protein Protocols, Vol. 2: Methods and Techniques*
Edited by: H. J. Vogel 

actions in the cell. When interpreting the fluorescent signals, it has to be considered that both the concentration and the liganded state of calmodulin may change simultaneously. Vertebrate calmodulins can be modified most readily on lysine residues. Several nonfluorescent lysine reagents have been attached to calmodulin and, in most cases, the reagents predominantly modified $Lys_{75}$ ***(2)***. The higher reactivity of $Lys_{75}$ as opposed to the other lysine residues in the protein, originates in its lower $pK_a$ value as determined by NMR spectroscopy ***(3)***. Here, we have used the isomeric fluorescein dichlorotriazine, 5-DTAF, which readily modifies amines in proteins. In this chapter, we use the example of $Lys_{75}$-labeled FL-calmodulin, which has already been used for imaging in living cells. We intend to provide a comprehensive approach including the synthesis and analysis of the fluorescent calmodulin and its application into living cells. We close with demonstrating how the activation state of calmodulin can be monitored in living cells by confocal imaging.

## 2. Materials

### 2.1. Synthesis of FL-Calmodulin

1. 4 m*M* 5-DTAF (D-16, Molecular Probes), stock solution dissolved in dimethyl formamide (DMF).
2. 60 μ*M* pig brain calmodulin ***(4)*** in 200 m*M* Tris-HCl (pH 8.5) containing 20 m*M* $CaCl_2$.
3. PD10 Columns (Pharmacia Biotech).
4. Solvent A: 0.1% solution trifluoroacetic acid (TFA) (HPLC grade)/$H_2O$.
5. Solvent B: 0.082% TFA solution/acetonitrile (HPLC grade).
6. Semipreparative Vydac C18 reverse-phase column 10 × 250 mm (Hichrom Ltd., UK).
7. 1 *M* Tris-HCl buffer, pH 7.5.

### 2.2. Characterization of $Lys_{75}$ Singly Labeled FL-Calmodulin

1. Freeze dried, desalted, singly labeled $Lys_{75}$ FL-calmodulin
2. Digestion mixture containing 100 m*M* $NH_4HCO_3$, pH 9.0, 2 m*M* EGTA, and 10 μg/mL trypsin, TPCK treated (Sigma, Aldrich, UK).
3. Solvent A: 0.1% solution trifluoroacetic acid (TFA) (HPLC grade)/$H_2O$.
4. Solvent B: 0.082% TFA solution/acetonitrile (HPLC grade).
5. Semipreparative Vydac C18 reverse-phase column 10 × 250 mm (Hichrom Ltd.).
6. Electrospray mass spectrometer.

### 2.3. Other $NH_2$-Reactive Fluorescent Probes

1. Dansyl chloride.
2. TA-Cl [2,4-dichloro-6-(4-*N*,*N*-diethylaminophenyl)-1,3,5-triazine].
3. Texas Red.
4. Cy5 [N-(N-hydoxy-succinimidyl-carboxypentyl)-$N^1$-(ethyl)-indodicarbocyamine-5,$5^1$-disulforate].

### 2.4. Materials for Electroporation

1. 250 mL stock solution containing: 135 m*M* NaCl, 5 m*M* KCl, 20 m*M* HEPES (pH 7.4 adjusted with NaOH), 2 m*M* $MgSO_4$, 10 m*M* glucose (add fresh).
2. 50 mL stock solution without $Ca^{2+}$ (solution A).
3. 100 mL stock solution containing 2 m*M* $CaCl_2$ (solution B).
4. 50 mL stock solution containing 1 m*M* EGTA (poration buffer, solution C).
5. 100 μL stock solution containing 1 m*M* EGTA, 0.5 μ*M* Tetrodotoxin (TTX), 2 m*M* FL-calmodulin (injection solution, solution D).
6. 5 mL disposable syringes (BDH/Merck) with sterile Puradiscs™ (Whatman, 25 mm diameter, 0.2-μm pore size).
7. 60-mm disposable, sterile, tissue-culture dishes (Costar).
8. Cover slips containing freshly prepared DRG cells.

### 2.5. Materials for Microinjection

1. Early *Lytechinus pictus* embryos in artificial sea water (410 m*M* NaCl, 39 m*M* $MgCl_2$, 15 m*M* $MgSO_4$, 2.5 m*M* $NaHCO_3$, 10 m*M* $CaCl_2$, 10 m*M* KCl, and 1 m*M* EDTA; pH is adjusted to 8.0 and osmolarity to 950–1000 mosmol).
2. Stock solutions of 5 m*M* singly labeled FL-calmodulin, 5 m*M* FL-Dextran, and 10 m*M* TA-calmodulin. All reagents were dissolved in an injection buffer containing 0.5 *M* KCl, 20 m*M* PIPES, pH 7.2, and 100 μ*M* EGTA.
3. Borosilicate glass micropipets (Clark Electromedical Instruments).
4. High-pressure injector system equipped with a hydraulic manipulator (Narishige Instruments).

### 2.6. Materials for Microscopy and Imaging

1. Bio-Rad MRC 1000 UV confocal microscope.
2. Argon UV laser lines of 351 and 363 (for TA-calmodulin) and 488-nm argon laser (for FL-calmodulin).
3. Emitted light is directed through a 450-nm dichroic mirror into separate detectors where a 405-nm, 35-nm FWH*M* bandpass (to detect TA-calmodulin) and a 530-nm longpass (to detect FL-calmodulin) filter were used to create images.
4. Calmodulin or Dextran.
5. Freshly shed eggs of *Lytechinus pictus* microinjected with fluorescent calmodulin or Dextran.

## 3. Methods

### 3.1. Preparation of FL-Calmodulin

1. Calmodulin (2.5 mg) in 200 m*M* Tris/HCl (pH 8.5) containing 20 m*M* $CaCl_2$ is treated with 150 μ*M* 5-DTAF (from 4 m*M* 5-DTAF in DMF). Labeling is routinely carried out in 20 m*M* $CaCl_2$ to maximize the rate and the specificity of the reaction of the calmodulin with 5-DTAF ***(5)***. At this pH and $Ca^{2+}$ concentration, $Lys_{75}$ is labeled highly selectively with 5-DTAF.
2. The solution is allowed to react at 22°C in the dark for 40 min.

3. The reaction can then be terminated by gel-filtration on a Sephadex PD10 column equilibrated in $H_2O$. Excess insoluble and soluble reagent is removed on addition of 2.5 mL of the reaction mixture to the column and elution of 3.5 mL with $H_2O$ (*see* **Note 1**).
4. Singly labeled FL-calmodulin (calmodulin labeled with 5-DTAF on $Lys_{75}$) is resolved from unlabeled calmodulin and doubly labeled FL-calmodulin (calmodulin labeled with 5-DTAF on $Lys_{75}$ and $Lys_{148}$) using HPLC with a Vydac reverse-phase $C_{18}$ column (10 × 250 mm). For analytical purposes, an aliquot (10 μL) is chromatographed at a flow rate of 2.5 mL/min with a linear gradient from 70% solvent A — 30% solvent B to 30% solvent A — 70% solvent B over 40 min. Absorption is measured at 215 nm and fluorescence is monitored at 450 nm (excitation) and 526 nm (emission). *See* **ref. *6*** for HPLC methodology.
5. The HPLC analysis shown in **Fig. 1**, should show three main absorption peaks in the following order:
   a. Unlabeled calmodulin — absorption peak with no associated fluorescence
   b. Singly labeled calmodulin — absorption peak with associated fluorescent peak.
   c. Doubly labeled calmodulin — absorption peak with associated fluorescent peak.
6. The preparative procedure follows the aforementioned method where singly labeled FL-calmodulin is purified in several batches. The UV-absorbing peaks corresponding to calmodulin, singly labeled FL-calmodulin, and doubly labeled FL-calmodulin are collected, pooled, and freeze dried.
7. The freeze-dried product is desalted by dissolving it in Tris-HCl, pH 7.5 buffer and passing the solution through an $H_2O$-equilibrated PD10 column. Again, 2.5 mL of the solution is applied and 3.5 mL is eluted with $H_2O$. This solution is then freeze dried.
8. Protein molecular weights were determined by electrospray ionization mass spectrometry ***(7,8)*** on a Platform single-quadrupole mass spectrometer (Micromass, UK). Proteins were desalted prior to analysis using a 2 mm × 2 cm column (Upchurch Scientific, Oak Harbor, WA) slurry packed with poros R2 (Perseptive Biosystems, Framingham, MA) and fitted across ports 1 and 4 of a Rheodyne 7000 valve. 100–200 pmol of protein diluted in 10% acetonitrile, 0.1% formic acid buffer were loaded onto the column via port 5 and were desalted with 250–1000 μL of the same buffer depending on the initial salt concentration. A 130-A syringe pump (Perkin Elmer) running 70% acetonitrile, 0.1% formic acid at 10 μL/min was connected to port 2. After desalting of the protein, the Rheodyne was switched to connect ports 1–2 and 3–4 (with port 4 connected to the mass spectrometer) and thus protein was eluted off the column into the mass spectrometer. The mass spectrometer was operated at an electrospray voltage of 3.5 kV, a cone voltage of 30 V, and was calibrated using myoglobin. Electrospray mass spectrometry of singly and doubly labeled FL-calmodulin gave a series of peaks that correspond to protein molecules with varying net charges *z*.

   **Figure 2A** shows a mass-to-charge ratio *m/z* of each of the major peaks, the average mass is 17254.7 (± 5) Da. This mass represents calmodulin (16791.4 Da) with

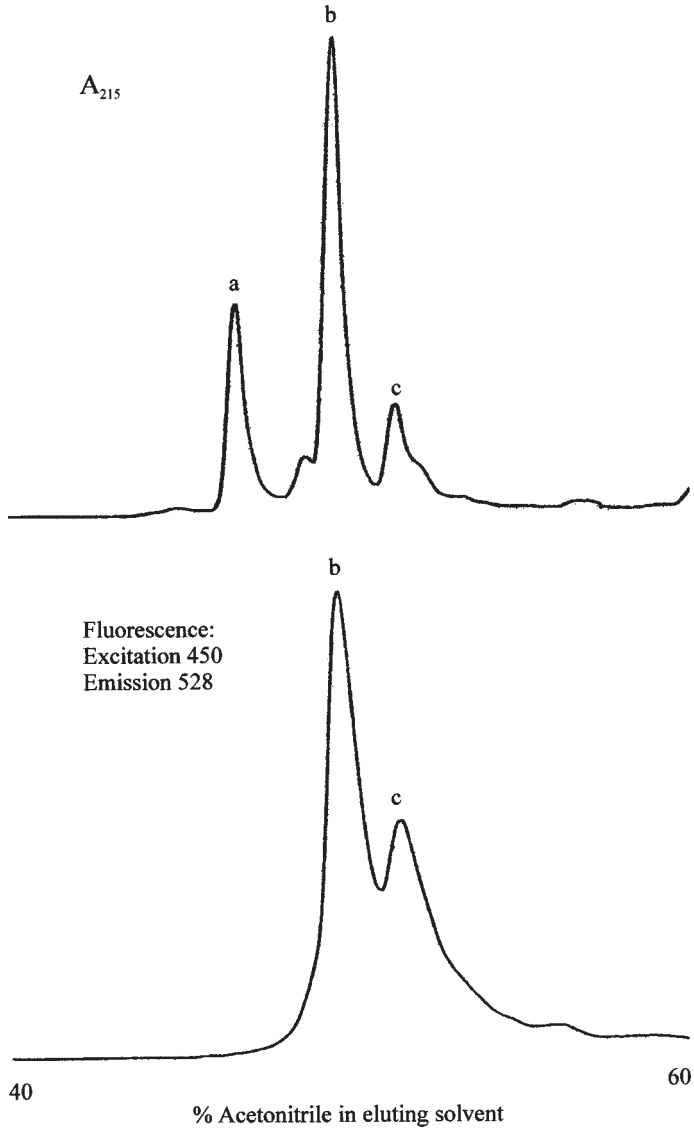


Fig. 1. Purification of singly labeled FL-calmodulin. The FL-calmodulin reaction mixture after PD-10 filtration was analyzed by HPLC on a semipreparative Vydac C18 reverse phase column. The mixture shows three peaks in the following order: **(a)** unlabeled calmodulin; **(b)** singly labeled FL-calmodulin; **(c)** doubly labeled FL-calmodulin. The illustrated section of the chromatograms represent the 40–60% acetonitrile in the eluting solvent gradient. Absorbance (*top panel*) was measured at 215 nm and fluorescence (*bottom panel*) was measured with transmission peaks of 450 nm (excitation) and 528 nm (emission).

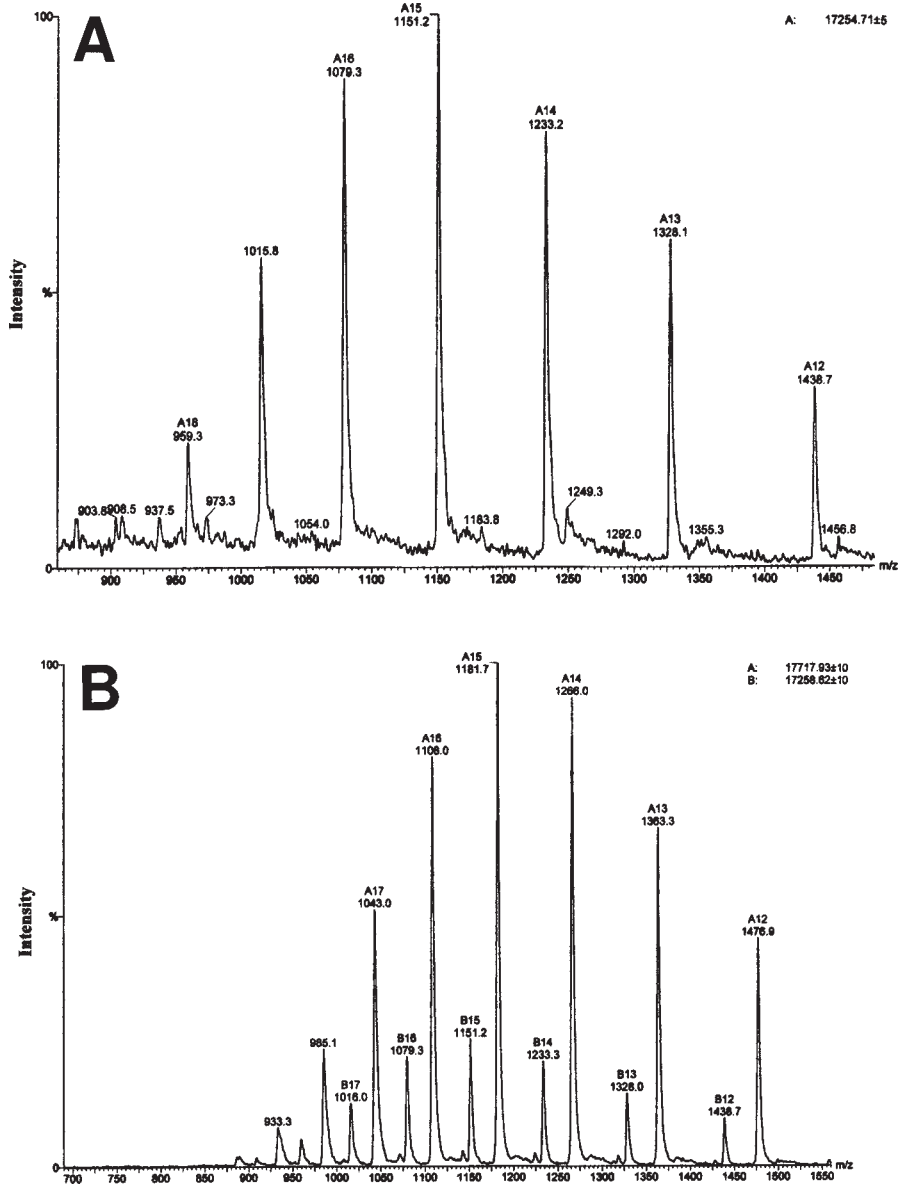


Fig. 2. Electrospray mass spectrometry of undigested singly labeled and doubly labeled FL-calmodulin gave a series of peaks that correspond to protein molecules with varying net charges z. 100–200 pmol of protein were diluted in 10% acetonitrile 0.1% formic acid solvent. **(A)** shows a mass-to-charge ratio *m/z* of each of the major peaks, the average mass is 17254.7 (± 5) Da. This mass represents calmodulin (16791.4 Da) bound to 5-DTAF (458.5 Da) at a single site. **(B)** shows two series of peaks, one represents calmodulin bound to 5-DTAF at two sites (17718 [± 10] Da) and the other series shows contamination by singly labeled FL-calmodulin (17258.62 [± 10] Da).

5-DTAF (458.5 Da) attached at a single site. **Figure 2B** shows two series of peaks, one represents calmodulin with 5-DTAF attached to two sites (17718 [± 10] Da) and the other series shows contamination of singly labeled FL-calmodulin (17258.62 (± 10) Da).

### 3.2. Characterization of $Lys_{75}$ Singly Labeled FL-Calmodulin

1. Tryptic Digestion: Unlabeled calmodulin, singly labeled and doubly labeled FL-calmodulin produce characteristic peptide maps when digested with the proteolytic enzyme trypsin, which catalyzes the hydolysis of lysyl and arginyl peptide bonds ***(9)***. These peptide maps allow a fast and simple way of determining the extent of modification of calmodulin, which is important because for proper imaging results we require pure singly labeled FL-calmodulin. Analysis of these digests using HPLC will produce a number of peaks each of which will correspond to an individual peptide fragment.
2. The digestion mixture containing 100 m*M* $NH_4HCO_3$, pH 9.0, 2 m*M* EGTA, and 10 μg/mL trypsin is added to the freeze-dried, desalted, unlabeled calmodulin, singly labeled FL-calmodulin (or the doubly labeled FL-calmodulin).
3. Digestions are carried out at 37°C. Labeled FL-calmodulin is substantially more resistant to tryptic cleavage than calmodulin and usually requires overnight digestion. Unlabeled calmodulin is completely cleaved in 2 h.
4. Digestion is terminated by the addition of 5 vol of 0.1% TFA.
5. Tryptic digests can then be analyzed by reverse phase HPLC using a Vydac reverse phase $C_{18}$ column (10 × 250 mm). For analytical purposes, a 50-μL aliquot is chromatographed at a flow rate of 2.5 mL/min with a linear gradient from 95% solvent A — 5% solvent B to 35% solvent A — 65% solvent B over 60 min. Absorption is measured at 215 nm and fluorescence is monitored as before at 528 nm (emission). The chromatogram of a typical tryptic digest of singly labeled FL-calmodulin can be seen in **Fig. 3**. **Figure 4** shows the amino acid sequences of the predicted tryptic peptide fragments of calmodulin.
6. If the procedure is performed on the preparative procedure scale, then follow the aforementioned method where peptides are collected corresponding to UV-absorbing peaks. These peptides are then freeze-dried ready for analysis by mass spectrometry.
7. Electrospray/nanospray MS: Characterization of peptide peaks is done by electrospray and nanospray mass spectrometry. For a more-detailed description for materials and methodology on nanospray mass spectrometry, refer to ***(10)***. Average masses of peptide fragments isolated by HPLC were discovered by using electrospray and nanospray mass spectrometry. **Table 1** relates the HPLC peaks obtained to the various labeled and unlabeled peptide fragments of calmodulin.
8. From the HPLC trace of the trypic digest of singly labeled FL-calmodulin seen in **Fig. 3**, there are two main fluorescent peaks, peaks 7 and 9. Analysis of peak 7, as seen in **Fig. 5**, by electrospray mass spectrometry gave two average masses. The first mass relates to peptide contamination from the preceeding peak, whereas the second mass originates from peptide fragment 75–86 with the 5-DTAF probe

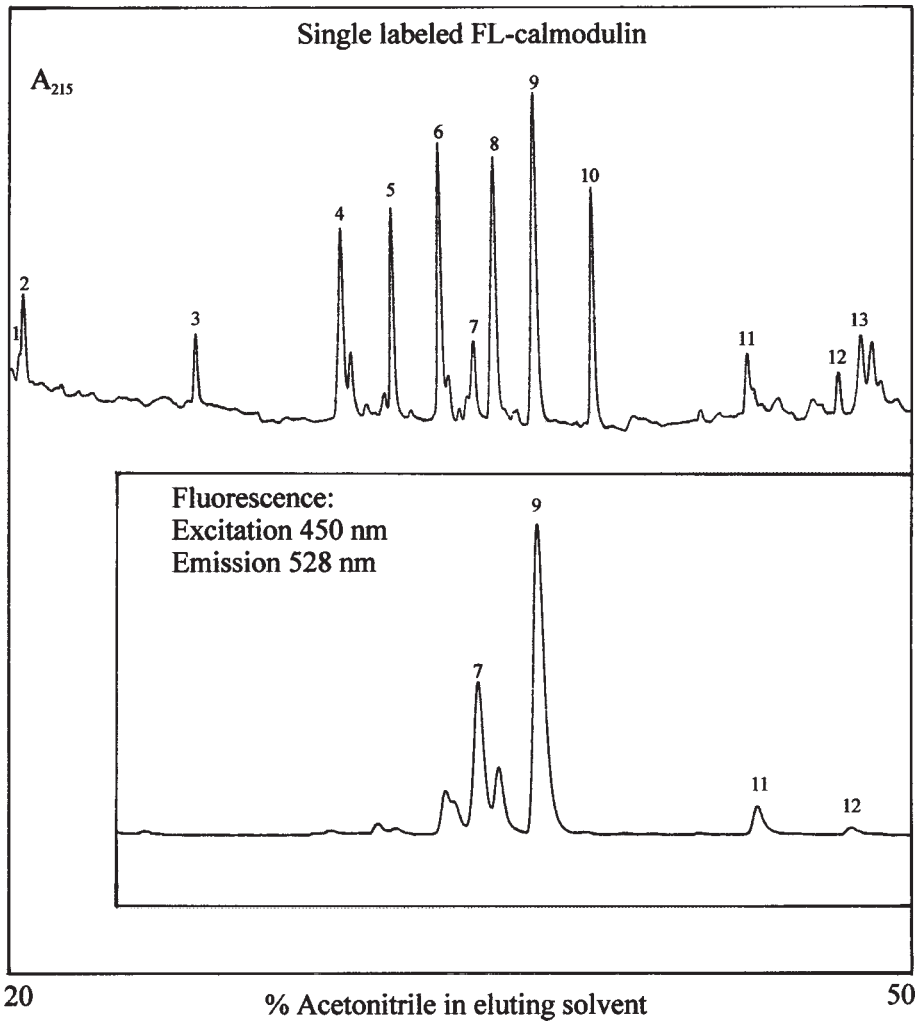


Fig. 3. Peptides from a pH 9.0, 20 h tryptic digest of singly labeled FL-calmodulin obtained from the 20–50% acetonitrile part of the gradient on a Vydac HPLC column. The chromatograms show absorbance at 215 nm and fluorescence at transmission peaks of 450 nm (excitation) and 528 nm (emission). Peaks 7 and 9 show the majority of the fluorescence, whereas peaks 11 and 12 show slight fluorescent properties. This represents the 5-DTAF fluorophore bound to lysine residues on the peptide fragments. Other fluorescent peaks in the digest were not identified. In doubly labeled FL-calmodulin a new absorption and fluorescent peak appears at position 12 (residues 127–148), which corresponds to the second labeled site ($Lys_{148}$) not seen in singly labeled FL-calmodulin. Also, there is a reduced absorption peak at position 10 because of the labeling of this same peptide fragment in position 12.

**Human Calmodulin**
**Mass = 16791.4**
**N-Terminus = $CH_3CO$**
**C-Terminus = OH**

| | T1 | | T2 | | T3 | | T4 | |
|---|---|---|---|---|---|---|---|---|
| [ADQLT | EEQIA | EFK][EA | FSLFD | K][DGDG | TITTK] | [ELGTV | MR] | |
| 1 | 6 | 11 | 16 | 21 | 26 | 31 | 36 | |

| | | | T5 | | | | |
|---|---|---|---|---|---|---|---|
| [SLG | QNPTE | AELQD | MINEV | DADGN | GTIDF | PEFLT | MMAR[K | |
| | 41 | 46 | 51 | 56 | 61 | 66 | 71 | |

| | T8 | | T9 | T10 | | T11 | | |
|---|---|---|---|---|---|---|---|---|
| MK]DTD | SEEEI | R][EAFR] | | [VFDK][D | GNGYI | SAAEL | R][HMVT | |
| 76 | 81 | 86 | | 91 | 96 | 101 | 106 | |

| | T12 | | | | T13 | | | |
|---|---|---|---|---|---|---|---|---|
| NLGEK | LTDEE | VDEMI | R][EADI | DGDGQ | VNYEE | FVQMM | TAK | |
| 111 | 116 | 121 | 126 | 131 | 136 | 141 | 146 | |

Fig. 4. Amino acid sequence of human calmodulin showing peptide fragments expected from a typical tryptic digest, which catalyzes the hydrolysis of lysyl and arginyl peptide bonds. $Lys_{75}$ and $Lys_{148}$ are shown in bold to show that these were the two main 5-DTAF labeling sites in doubly labeled FL-calmodulin and $Lys_{75}$ only in singly labeled FL-calmodulin.

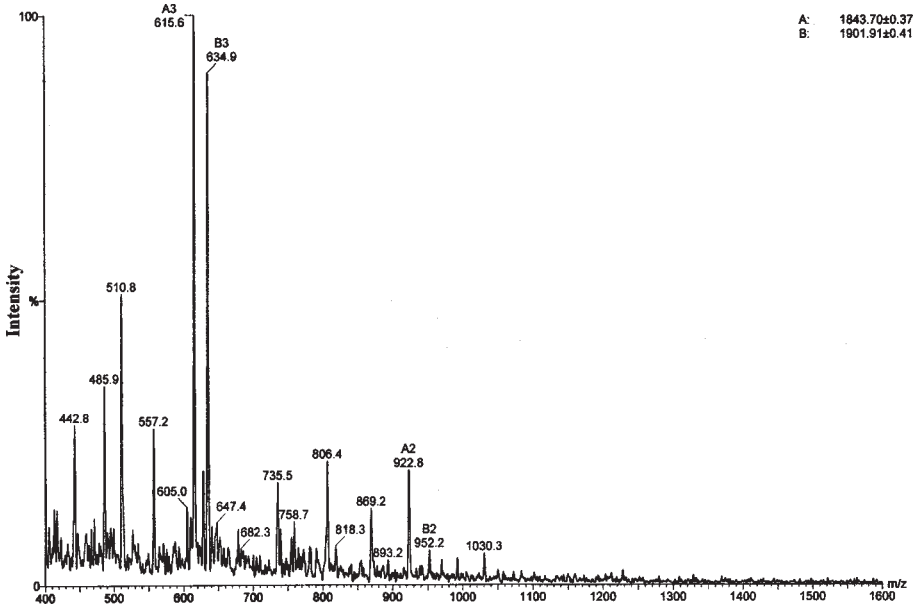


Fig. 5. Electrospray mass spectrum of the desalted, freeze dried peptide from peak 7 isolated by HPLC dissolved in 10% acetonitrile 0.1% formic acid solvent. The first mass (1843.70 (± 0.37) Da) relates to peptide contamination from the preceeding peak, whereas the second mass (1901.91 (± 0.41) Da) relates to peptide fragment 75–86 with the 5-DTAF probe intramolecularly cyclized between two amino groups by substitution of both chlorines.

**Table 1**
**Peptide Fragments Isolated by HPLC on a Vydac Semipreparative Column; Peptides were Analyzed by Electrospray and Nanospray Mass Spectrometry**

| Peak | Observed Average Mass | Sequence | Modified Residues | Expected Average Mass |
|---|---|---|---|---|
| 1 | 1092.5 | T8 (77–86) | | 1093.1 |
| 2 | 521.5 | T9 (87–90) | | 521.6 |
| 3 | 803.9 | T4 (31–37) | | 804.9 |
| 4 | 1753.7 | T10–11 (91–106) | | 1753.9 |
| 5 | 1562.8 | T1 (1–13) | Acetyl Group on $Alanine_1$ | 1562.8 |
| 6 | 1843.9 | T2–T3 (14–30) | | 1843.9 |
| 7 | 1901.9 | 75–86 | *5-DTAF cyclized | 1902 |
| 8 | 2400.5 | T12 (107–126) | $(CH_3)_3$ on $Lysine_{115}$ | 2402 |
| 9 | 864.1 | 75–77 | 5-DTAF on $Lysine_{75}$ | 863.6 |
| 10 | 2489.1 | T13 (127–148) | | 2490 |
| 11 | 2302.5 | T2–T3 (14–30) | 5-DTAF on $Lysine_{21}$ | 2301.9 |
| 12[a] | 2951 (127–148) | T13 on $Lysine_{148}$ | 5-DTAF | 2948.7 |
| 13 | 4070.7 | T5 (37–74) | | 4071.5 |

[a]Doubly labeled only

intramolecularly cyclized between two amino groups by substitution of both chlorines (*see* **Fig. 6**). Analysis of peak 9 by nanospray mass spectrometry shown in **Fig. 7A**, shows a monoisotopic mass of 864.1, which identifies peptide fragment 75–77 with a 5-DTAF label on one of the two lysines. In **Fig. 7B**, we can see the entire amino acid sequence and which lysine residue is labeled with 5-DTAF. In this case, the label is on $Lys_{75}$ and not on $Lys_{77}$ because the b ions (fragmentation occurring at the amide backbone to produce acylium ions) and the y" ions (*C*-terminal ions involving hydrogen rearrangement) would produce different masses as outlined in **Fig. 8** (*see* **Note 2**). The reaction mechanism that gives these two different peptides is shown in **Fig. 6**. As aforementioned, in the

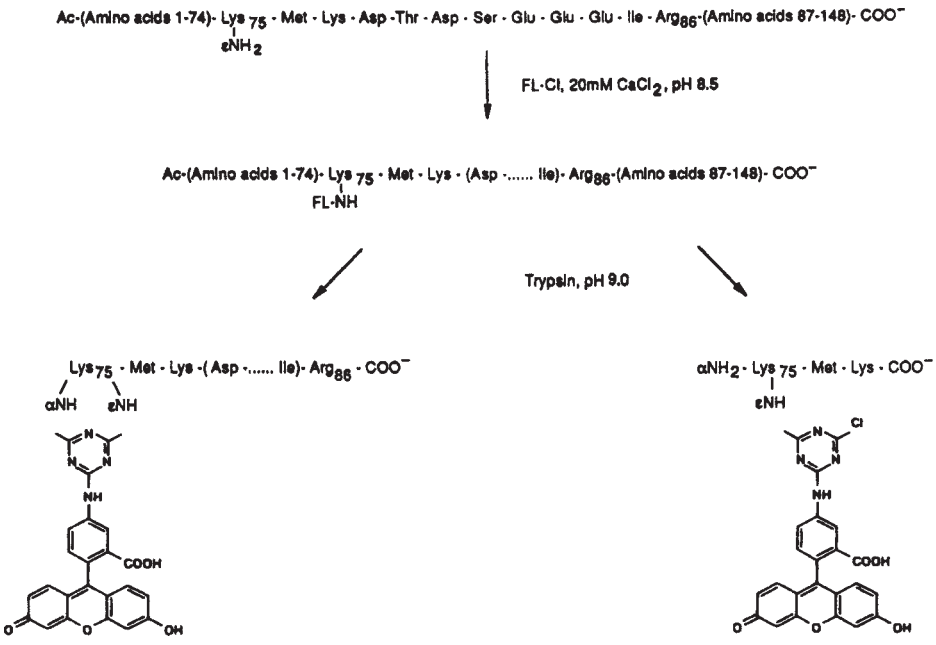


Fig. 6. Reaction mechanism for the synthesis of the two major fluorescent peaks (7 and 9) in singly labeled FL-calmodulin. In the longer peptide (residues 75–86), the probe is intramolecularly cyclized between two amino groups by substitution of both chlorines and as reaction conditions favored $Lys_{75}$ labeling, it is expected that the probe is cyclized on this residue rather than $Lys_{77}$. In the case of the short peptide (residues 75–77), 5-DTAF is bound only to $Lys_{75}$ by substitution of only one of the chlorines and so in this case is not cyclized.

longer peptide (residues 75–86), the probe is intramolecularly cyclized between two amino groups by substitution of both chlorines and as reaction conditions favored $Lys_{75}$ labeling, it is expected that the probe is cyclized on this residue rather than $Lys_{77}$. In the case of the short peptide (residues 75–77), 5-DTAF is bound only to $Lys_{75}$ by substitution of only one of the chlorines and so in this case is not cyclized.

9. The materials and methodology used for characterization of $Lys_{75}$ and $Lys_{148}$ doubly labeled FL-calmodulin are identical as described for singly labeled FL-calmodulin. However, the tryptic digests of doubly labeled FL-calmodulin and singly labeled FL-calmodulin show two important differences. A new absorption and fluorescent peak appears at position 12, which corresponds to the second labeled site not seen in singly labeled FL-calmodulin. Also, a reduced absorption peak at position 10 is seen since this peptide fragment becomes labeled, reappears as fluorescent peptide peak 12 (*see* **Fig. 3**). **Figures 9** and **10**

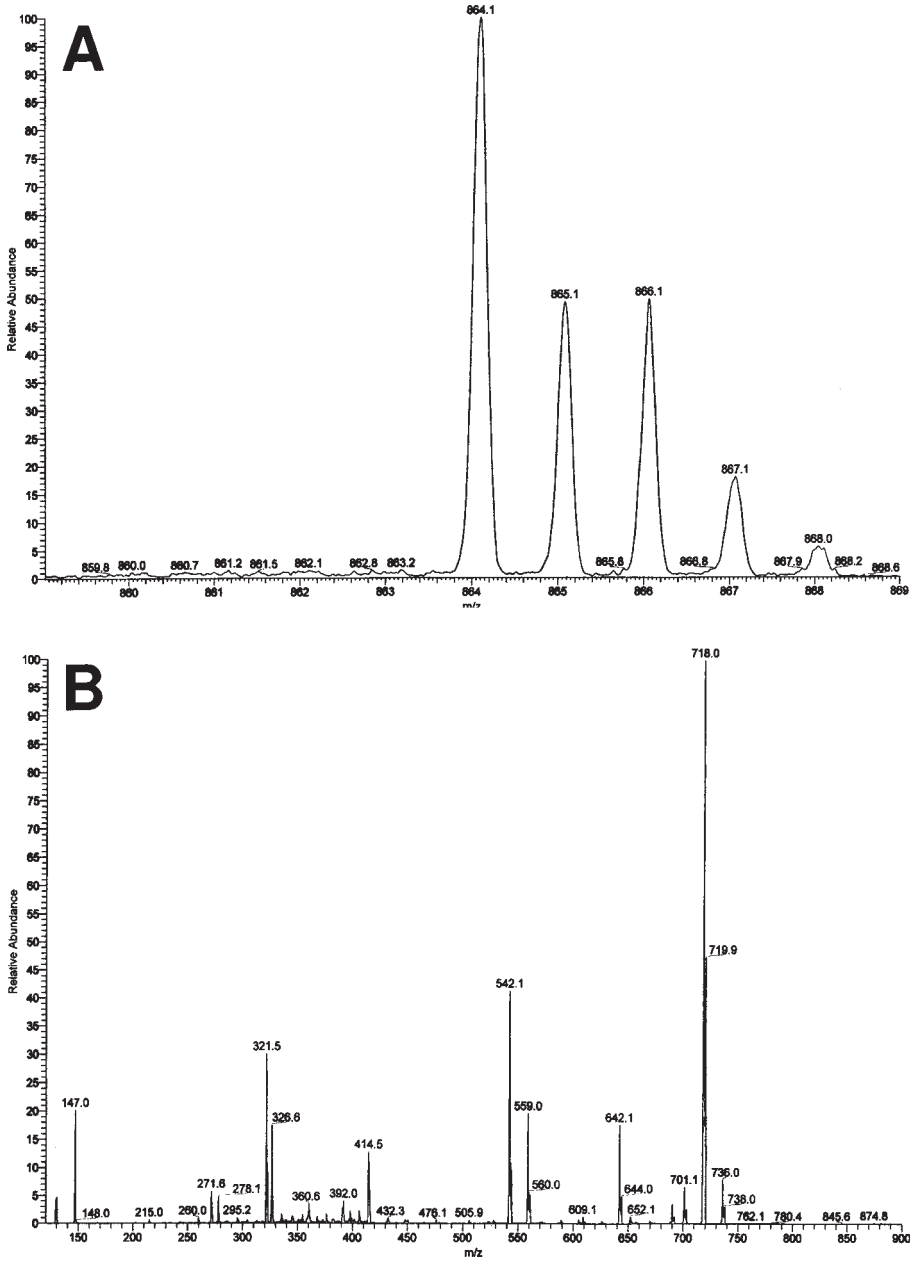


Fig. 7. Nanospray mass spectra of the desalted, freeze-dried peptide from peak 9**.** **(A)**, shows a monoisotopic protonated ion of 864.1 Da which identifies peptide fragment 75–77 with a 5-DTAF label on one of the two lysines. **(B)** shows a fragmentation spectrum of the consecutive series of b and y" ions whose differences correspond to the residue masses of amino acids. In this case, the b ions (fragmentation occurring at the amide backbone to produce acylium ions) and the y" ions (*C*-terminal ions involving hydrogen rearrangement) gave masses that would only be seen if 5-DTAF was on $Lys_{75}$.

**Average Mass = 864.3792**
**Monoisotopic Mass = 863.2827**
**Residues 75-77**
**N-Terminus = H**
**C-Terminus = OH**
**Fragment ions: Monoisotopic m/z ratios with 1 positive charge (s)**

**a) KDt = $Lys_{77}$ labeled with 5-DTAF** **hypothetical**

| | | | |
|---|---|---|---|
| b | 129.1 | 260.1 | - |
| | 1 | 2 | 3 |
| | Lys | Met | KDt |
| | 3 | 2 | 1 |
| y'' | - | 736.2 | 605.2 |

**b) KDt = $Lys_{75}$ labeled with 5-DTAF** **found**

| | | | |
|---|---|---|---|
| b | 587.1 | 718.2 | - |
| | 1 | 2 | 3 |
| | KDt | Met | Lys |
| | 3 | 2 | 1 |
| y'' | - | 278.2 | 147.1 |

Fig. 8. Consecutive series of b and y" ions whose differences correspond to the residue masses of amino acids. **(a)** shows a hypothetical case where the 5-DTAF is on $Lys_{77}$ and the series of masses that would be seen from a nanospray mass spectrum. However, in this case, the residue masses in **(b)** are seen from the nanospray mass spectrum, therefore, 5-DTAF is bound to $Lys_{75}$.

show that nanospray mass spectrometry identified *T*13 (residues 127–148) and $Lys_{148}$ as the second labeled site in doubly labeled FL-calmodulin (*see* **Note 3**).

10. In both singly and doubly labeled FL-calmodulin peak 11 was fluorescent and analysis by electrospray and nanospray mass spectrometry identified the peptide as fragment T2–T3 (residues 14–30) with some labeling on $Lys_{21}$ (*see* **Figs. 11** and **12**). Other minor fluorescent peaks which were not identified may represent a small amount of labeling on the other lysines or other peptide fragments containing $Lys_{75}$.

### 3.3. Labeling with Other $NH_2$-Reactive Fluorescent Probes

**Table 2** shows a number of other fluorescent probes that have been used to specifically label calmodulin, but have not been fully characterized. With each

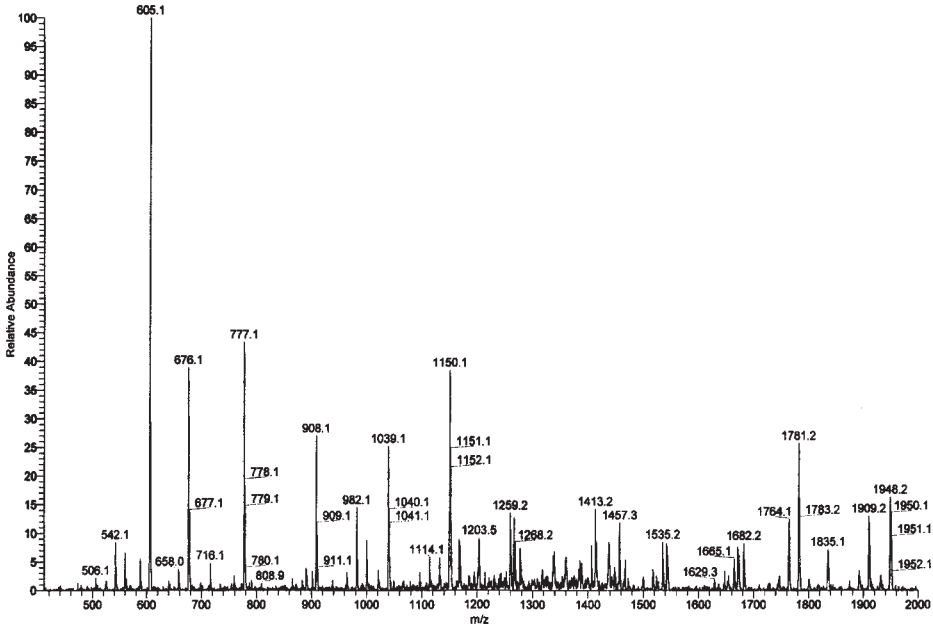


Fig. 9. Nanospray mass spectrometry identified peptide fragment T13 (residues 127–148) and $Lys_{148}$ as the second labeled site in doubly labeled FL-calmodulin. The spectrum shows a fragmentation spectrum of a consecutive series of b and y'' ions whose differences correspond to the residue masses of amino acids.

Average Mass = 2949.5257
Monoisotopic Mass = 2947.1145
Residues 127-148
N-Terminus = H
C-Terminus = OH
Fragment ions: Monoisotopic m/z ratios with 1 positive charge (s)

a) KDt = $Lys_{148}$ labeled with 5-DTAF

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| b | 130.1 | 201.1 | 316.1 | 429.2 | 544.2 | 601.2 | 716.3 | 773.3 | 901.4 | 1000 |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | **Glu** | **Ala** | **Asp** | **Ile** | **Asp** | **Gly** | **Asp** | **Gly** | **Gln** | **Val** |
| | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 |
| y'' | - | 2819 | 2748 | 2633 | 2520 | 2405 | 2348 | 2233 | 2176 | 2048 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| b | 1115 | 1278 | 1407 | 1536 | 1683 | 1782 | 1910 | 2041 | 2172 | 2273 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| | **Asn** | **Tyr** | **Glu** | **Glu** | **Phe** | **Val** | **Gln** | **Met** | **Met** | **Thr** |
| | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 |
| y'' | 1949 | 1835 | 1672 | 1543 | 1414 | 1266 | 1167 | 1039 | 908.3 | 777.2 |

| | | |
|---|---|---|
| b | 2344 | - |
| | 21 | 22 |
| | **Ala** | **KDt** |
| | 2 | 1 |
| y'' | 676.2 | 605.2 |

Fig. 10. Consecutive series of b and y'' fragment ions produced by nanospray mass spectrometry. Here, the C-terminal lysine ($Lys_{148}$) is labeled with 5-DTAF, not seen in singly labeled FL-calmodulin.

fluorescent probe, however, the stoichiometry of labeling in the main derivatized calmodulin was 1:1. The fluorescence excitation and emission bands of these probes span the near-UV and visible spectrum, providing the experimentalist with a range of options for confocal microscopy studies. Thus spatial distribution of calmodulin can be imaged using a fluorescently labeled calmodulin and a typically equipped confocal microscope.

## 3.4. Calmodulin Imaging: Delivery of Fluorescent Calmodulins into Living Cells

### 3.4.1. Electroporation

1. Electroporation has been used to momentarily permeabilize the cell membrane in the prepared dorsal root ganglion cells (DRGs), thereby allowing efficient entry of the fluorescently labeled calmodulins. The technology is based on the delivery of short and intense electrical pulses at an appropriate electrical-field strength to living cells, resulting in a transient and reversible alteration of the cell membrane. This allows the cell membrane to become more permeable to a large variety of hydrophilic molecules that are otherwise unable to diffuse through the cell membrane. This process has no effect on the molecule being inserted into the cell, it only facilitates its entry into the cell.
2. *See* **ref. *14*** for a description of microporator apparatus setup.
3. DRG medium is removed from a cover slip containing freshly prepared DRG cells using a 5-mL syringe.
4. 1.5 mL of solution B (stock solution containing 2 m*M* $Ca^{2+}$) is added on to the cover slip using a 5-mL sterile, disposable, syringe filtered through a sterile Puradisc (25-mm diameter, 0.2-μm pore size). This solution is removed and then a fresh 1.5-mL added and left immersing the cells for 2 min.
5. The above process is then repeated using solution A (stock solution containing no $Ca^{2+}$) so the cells are thoroughly washed and left immersed in this $Ca^{2+}$ free medium for 2 min.
6. This solution is removed and is replaced by 1.5 mL of solution C (poration buffer; stock solution containing 1 m*M* EGTA).
7. A microliter injector consisting of a Teflon tube with a wire tube is used to apply a small sample volume of 0.5–2 μL. In this case, the cells are exposed to 2 μL of solution D (injection solution; stock solution containing 1 m*M* EGTA, 0.5 μ*M* TTX, 2 m*M* singly-labeled FL-calmodulin).
8. Electroporation is performed with a pulse of 70 V and a duration of 30–100 ms applied to the cells. At this point, the cell membranes become permeable to the singly labeled FL-calmodulin.
9. After 30 s, the polarity is reversed and another 70 V pulse of 30–100 ms duration is fired at the cells.
10. After 1 min, the poration buffer is removed, the DRGs are overlaid with 1.5 mL solution B and left for 2 min.

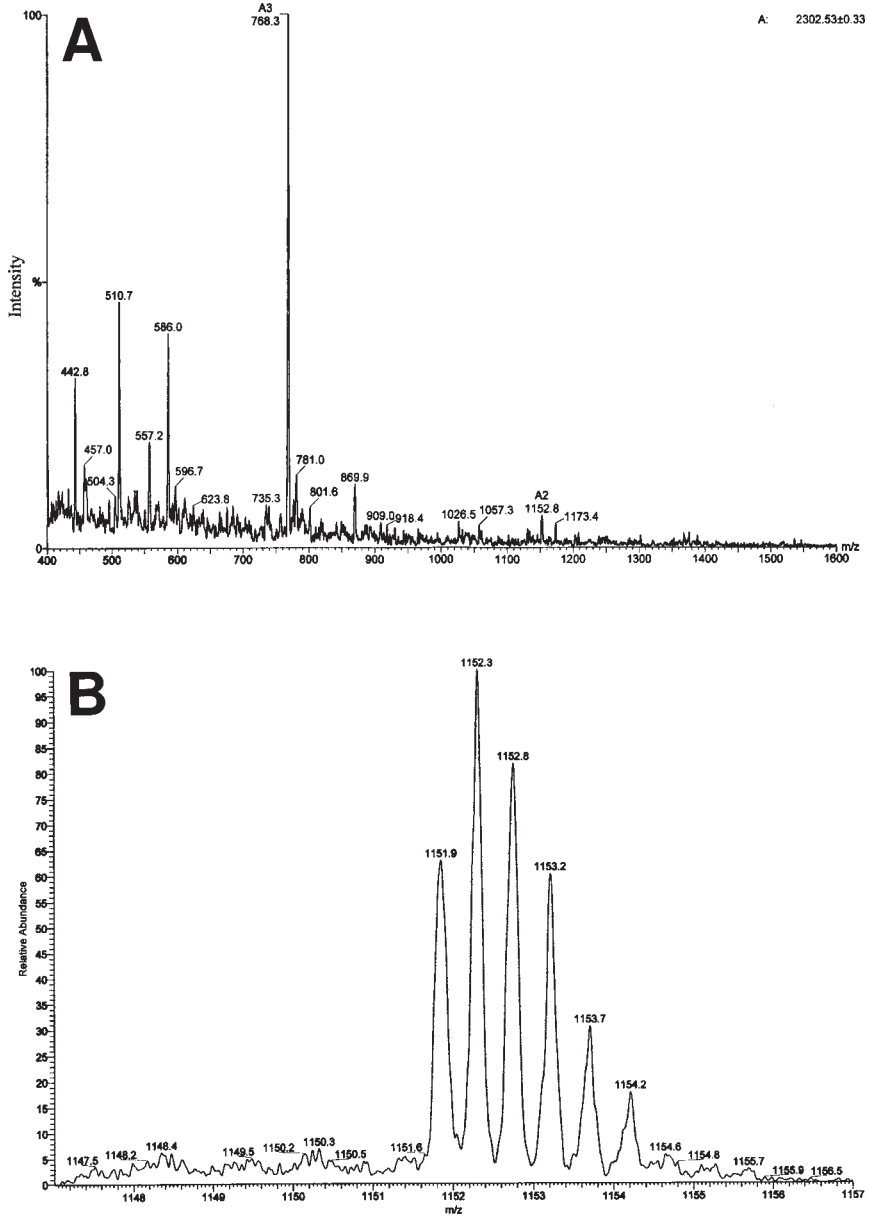


Fig. 11. Electrospray and nanospray mass spectrometry identified the peptide isolated by HPLC from absorption peak 11 as $T_2$-T3 (residues 14–30). This peak showed slight fluorescent properties suggesting that 5-DTAF labeled either $Lys_{21}$ or $Lys_{30}$. Fragmentation spectrum of a consecutive series of b and y" ions showed that 5-DTAF was bound to $Lys_{21}$. Panel B shows a 2+ ion giving a monoisotopic ion calculated $1151.9 \times 2 - 2$.

**Average Mass = 2303.8106**
**Monoisotopic Mass = 2301.9258**
**Residues 14-30**
**N-Terminus = H**
**C-Terminus = OH**
**Fragment ions: Monoisotopic m/z ratios with 1 positive charge (s)**

**KDt = $Lys_{21}$ labeled with 5-DTAF**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **b** | 130.1 | 201.1 | 348.2 | 435.2 | 548.3 | 695.3 | 810.4 | 1397 | 1512 | 1569 |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | **Glu** | **Ala** | **Phe** | **Ser** | **Leu** | **Phe** | **Asp** | **KDt** | **Asp** | **Gly** |
| | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 |
| **y''** | - | 2174 | 2103 | 1956 | 1869 | 1756 | 1609 | 1494 | 907.4 | 792.4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **b** | 1684 | 1741 | 1842 | 1955 | 2056 | 2157 | - |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| | **Asp** | **Gly** | **Thr** | **Ile** | **Thr** | **Thr** | **Lys** |
| | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| **y''** | 735.4 | 620.4 | 563.3 | 462.3 | 349.2 | 248.2 | 147.1 |

Fig. 12. Consecutive series of b and y'' fragment ions produced by nanospray mass spectrometry. The residue masses produced show that 5-DTAF is bound to $Lys_{21}$. This labeled peptide (residues 14–30) was found to be a minor product in both singly and doubly labeled FL-calmodulin.

**Table 2**
**Singly Labeled Fluorescent Calmodulins**

| Fluorescent Probe Bound to Calmodulin | Labeled Residue | Excitation nm | Emission nm |
|---|---|---|---|
| FL-calmodulin | $Lys_{75}$ | 488 | 528 |
| TA-calmodulin | $Lys_{75}$ | 365 | 415 |
| Cy5-calmodulin | n.d. | 649 | 662 |
| Texas Red-calmodulin | n.d. | 589 | 612 |

n.d. = not determined

11. At this point, the $Ca^{2+}$ containing medium is removed and the cells are reimmersed in 1.5-mL of DRG medium.
12. DRGs loaded with singly labeled FL-calmodulin can now be imaged by confocal microscopy.

### *3.4.2. Microinjection*

1. Freshly shed eggs of *Lytechinus pictus* microinjected with fluorescent freshly shed eggs of *Lytechinus pictus* are microinjected with the following fluorophores to a final concentration of: 5 μ*M* FL-calmodulin or 5 μ*M* FL-dextran,

3 μ*M* FL-calmodulin, 10 μ*M* TA-calmodulin, 3 μ*M* FL-calmodulin, and 10 μ*M* TA-calmodulin. For egg preparation see **ref. *15***.

2. The reagents are dissolved in an injection solution containing 0.5 *M* KCl, 20 m*M* PIPES pH 7.2, and 100 μ*M* EGTA.
3. Pulses of 0.1% of the cell volume are delivered using a high-pressure injector system equipped with a Narishige manipulator.
4. Estimates of the volume of the injection pulse are measured through displacement of cytoplasm and calculated by the relationship $\text{Vol} = 4/3 \times \pi \times \text{radius}^3$. The volume of the sea urchin egg is calculated to be roughly 500 pL, so the final concentration injected can be estimated.

### 3.5. Confocal Microscopy and Imaging

Digital confocal imaging enables a quantitative analysis of areas of fluorescence within cells. Here we use this technique to examine the mechanism of localization of calmodulin during the first cell cycle of the sea urchin zygote. However, because the small size of the microtubules prevents application of this type of analysis to the association of calmodulin with microtubules, the extent of calmodulin targeting to the nucleus and the mitotic apparatus is analyzed. (*See* **ref. *16*** for a description on scanning confocal microscopy.) Calmodulin images are analyzed off-line using Leica Lasertechnik software. Analysis is performed after filtering the images independently with a low-pass filter and then dividing the calmodulin-activation-sensitive channel pixel-by-pixel by the insensitive channel. Resultant images are processed by individually measuring pixel intensity values.

#### 3.5.1. Imaging of Calmodulin Localization

1. FL-, Cy5-, and Texas Red-calmodulins report calmodulin localization in the cell. We use the example of FL-calmodulin to illustrate calmodulin localization during mitosis. FL-calmodulin is injected to a final concentration of 5 μ*M* before the eggs are fertilized. Prior to fertilization, the calmodulin is localized within the nucleus whereas after fertilization the zygote nucleus centers and the astral microtubule arrays form until nuclear envelope breakdown (NEB) and entry into mitosis. In this period, calmodulin is localized in the nucleus and along the microtubule arrays. After the breakdown of the nuclear envelope, calmodulin is localized mainly in the spindle poles and can also be visualized in the location of the chromosomes. Finally, calmodulin is present within the reformed nuclei of the daughter cells as the embryo cleaves. This sequence of events can be seen in **Fig. 13**.
2. To test if the localization of FL-calmodulin was caused by specific targeting of the protein, fluorescein conjugated to a 10,000 molecular weight Dextran is used as a control. This is also injected before fertilization to a concentration of 5 μ*M*. FL-Dextran reports distributions in the cytoplasmic water space and perhaps nonspecific binding. FL-dextran is mainly localized cytoplasmically, whereas FL-calmodulin binds specifically to the astral tubule array. Also, the localization

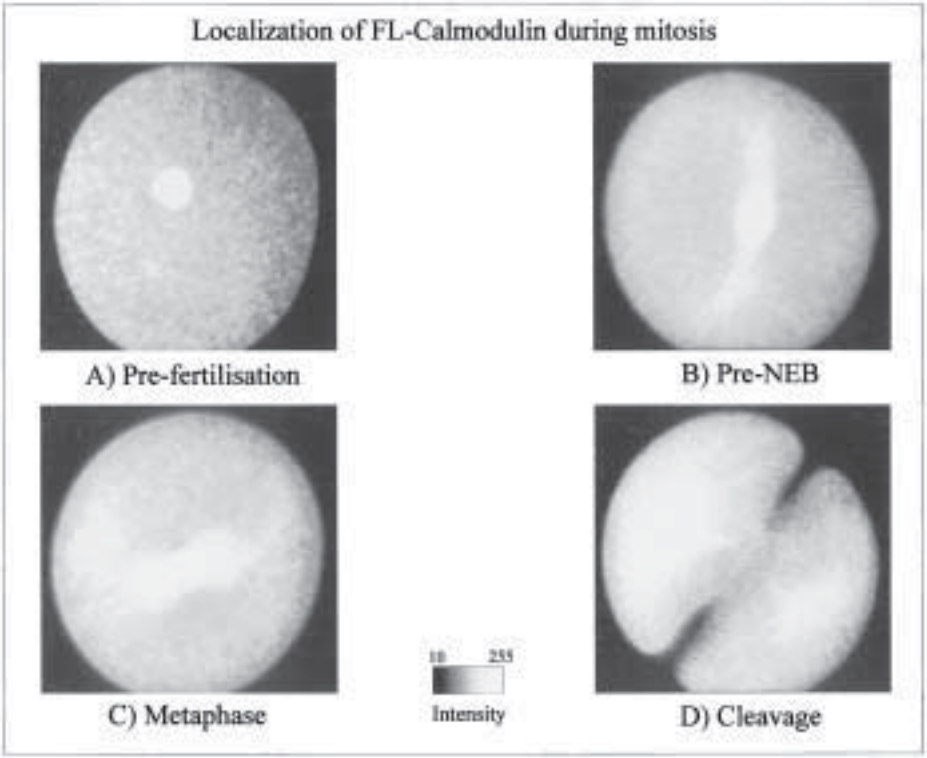


Fig. 13. Localization of DTAF-calmodulin during mitosis. 5 μ*M* of FL-calmodulin was microinjected and the cells were inspected on a Bio-Rad MRC 1000 UV confocal microscope with an argon laser line of 488 nm and 530 nm longpass (to detect FL-calmodulin) filter to create images. Prior to fertilization the calmodulin is localized within the nucleus whereas after fertilization the zygote nucleus centers and the astral microtubule arrays form until nuclear envelope breakdown (NEB) and entry into mitosis. In this period, calmodulin is localized in the nucleus and along the microtubule arrays. After the breakdown of the nuclear envelope, calmodulin is localized mainly in the spindle poles and can be visualized in the location of the chromosomes. Finally calmodulin is present within the reformed nuclei of the daughter cells as the embryo cleaves.

of FL-dextran is far less well defined, which suggests that binding of cellular structures is less specific. **Figures 13** and **14** show a comparison of the sequence of mitotic events between FL-calmodulin and FL-dextran.

### *3.5.2. Imaging of Calmodulin Activation*

1. As aforementioned, FL-calmodulin reports calmodulin localization in the cell. In contrast, TA-calmodulin fluorescence reports the interactions of calmodulin, as well as its concentration changes in the cell. In order to distinguish between

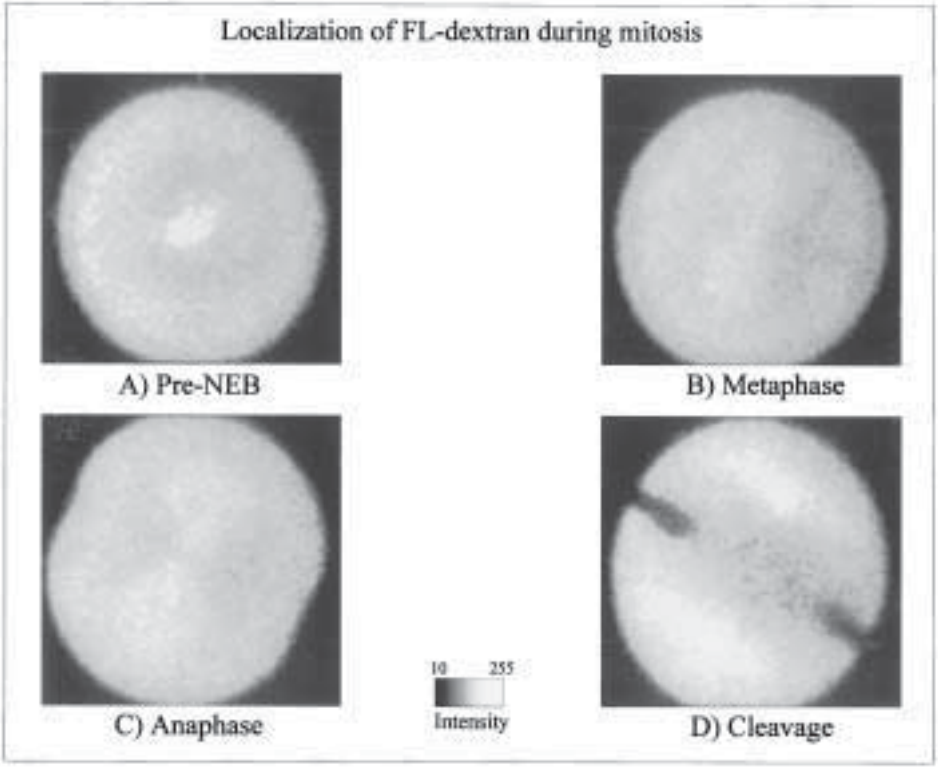


Fig. 14. Localization of FL-dextran calmodulin during mitosis. 5 μ*M* of FL-dextran was microinjected and the cells were inspected under the same conditions as for FL-calmodulin. FL-dextran reports distributions in the cytoplasmic water space and perhaps nonspecific binding. FL-dextran localizes cytoplasmically in comparison to FL-calmodulin, which binds specifically to the astral tubule array. Also, the localization of FL-dextran is far less well-defined, which suggests that binding of cellular structures is less specific.

the two events, TA- and FL-calmodulin are both applied to the same cell and their fluorescence emissions are scanned simultaneously. The fluorescence of FL-calmodulin microinjected into sea urchin eggs (final concentration 3 μ*M*) is relatively insensitive to $Ca^{2+}$ and target protein binding, but provides information on localization. **Figure 15** shows a sequence of mitotic transitions in a sea urchin egg microinjected with FL-calmodulin.

2. TA-calmodulin (final concentration of 10 μ*M*) microinjected into the sea urchin eggs shows a 10-fold rise in fluorescence intensity on $Ca^{2+}$ and target protein binding. **Figure 16** shows a sequence of mitotic transitions in a sea urchin egg microinjected with TA-calmodulin. Note the fluorescence intensity differences between FL-calmodulin and TA-calmodulin.

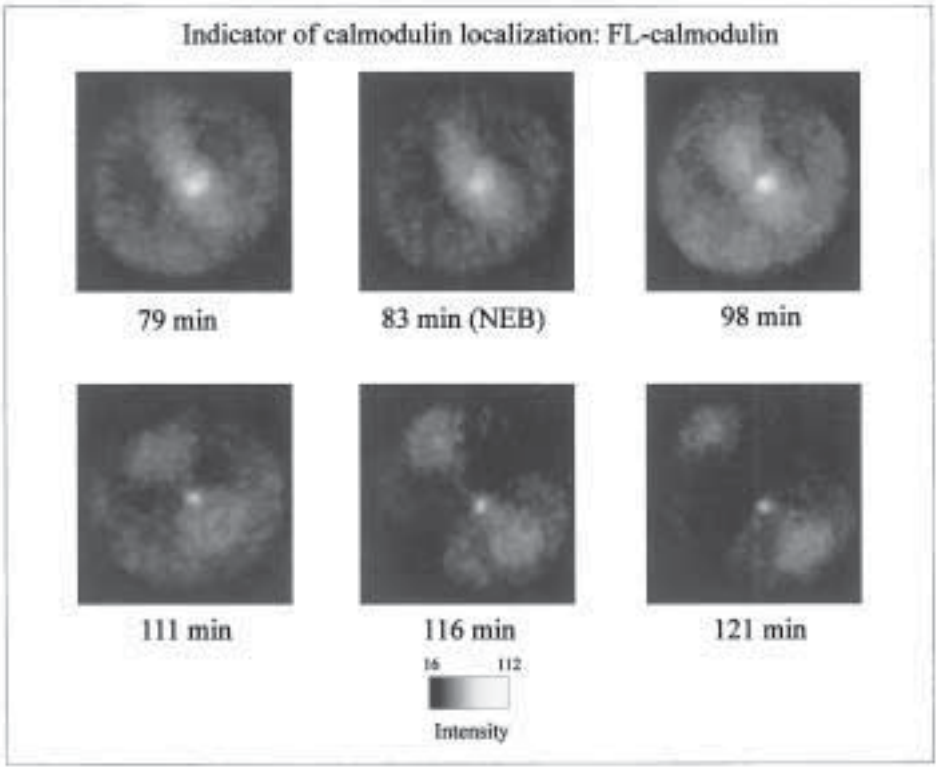


Fig. 15. Indicator of calmodulin localization using FL-calmodulin. The sequence shows mitotic events from 79–121 min where upon cell cleavage occurs. The fluorescence of FL-calmodulin (final concentration 3 μ*M*) microinjected into sea urchin eggs is relatively insensitive to $Ca^{2+}$ and target protein binding, but provides information on localization.

3. Simultaneous use of the $Ca^{2+}$-sensitive (TA-calmodulin) and insensitive (FL-calmodulin) derivative allow us to distinguish between $Ca^{2+}$ activation of calmodulin and local concentration changes of calmodulin (*see* **Fig. 17**).
4. To test that calmodulin activation is required for mitotic transitions Trp peptide (potent calmodulin inhibitor) ***(12,13)*** was injected prior to fertilization. Trp peptide blocks NEB and if injected postNEB blocks the metaphase-anaphase transition ***(17)***. This further suggests that $Ca^{2+}$-calmodulin-dependent processes are required for mitotic transitions.

## 4. Notes

1. The three important factors that must be observed while reacting 5-DTAF with calmodulin are pH, $Ca^{2+}$ (divalent cation) concentration, and no increase in reaction time. If these factors are not observed carefully, the ratio of singly labeled

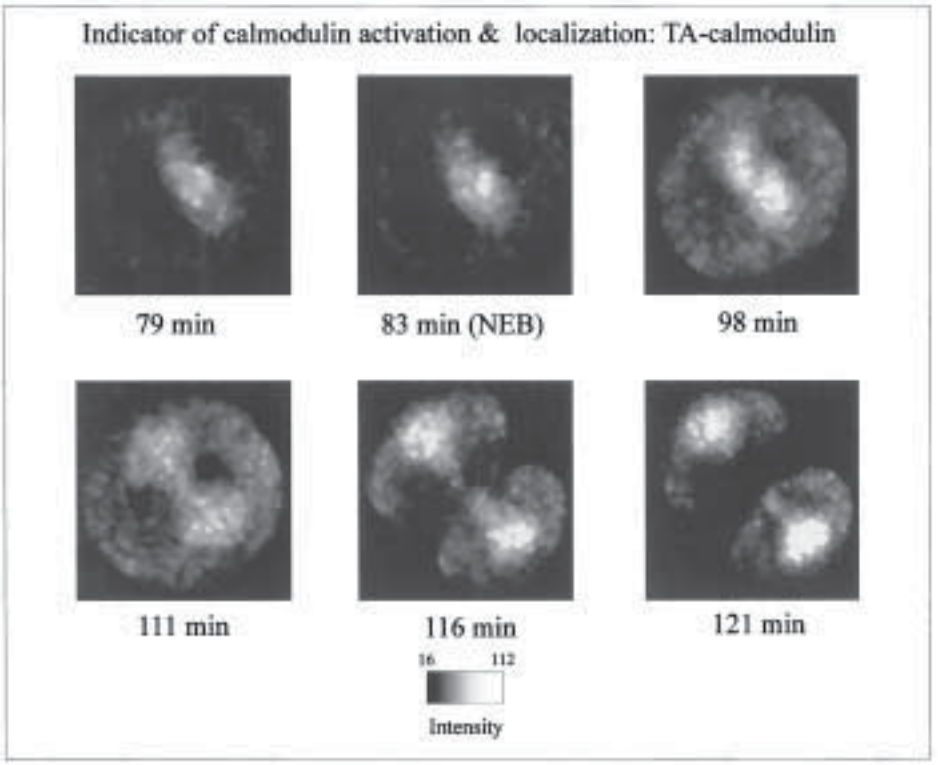


Fig. 16. Indicator of calmodulin activation and localization using TA-calmodulin. TA-calmodulin (final concentration of 10 μ*M*) microinjected into sea urchin eggs shows a 10-fold rise in fluorescence intensity on $Ca^{2+}$ and target protein binding. The sequence shows the mitotic events from 79–121 min.

FL-calmodulin (5-DTAF labeled on $Lys_{75}$ of calmodulin) to doubly labeled FL-calmodulin (5-DTAF labeled on $Lys_{75}$ and $Lys_{148}$ of calmodulin) will be reduced. We select singly-labeled FL-calmodulin so that precise measurements of fluorescence can be made during imaging.

2. Both $Lys_{75}$ and $Lys_{77}$ are located in a relatively exposed region of calmodulin. The fact that $Lys_{77}$ is not very reactive compared to $Lys_{75}$, originates from its relatively high $pK_a$ value ***(3)***.
3. In both singly and doubly labeled FL-calmodulin, peak 11 was fluorescent and analysis by electrospray and nanospray mass spectrometry identified the peptide as fragment T2–T3 (residues 14–30) with some labeling on $Lys_{21}$. Other minor fluorescent peaks that were not identified may represent a small amount of labeling on the other lysines or other peptide fragments containing the labeled $Lys_{75}$.
4. For imaging of spatial distribution of proteins, it is advantageous if the fluorophore covalently attached to calmodulin is not environmentally sensitive.

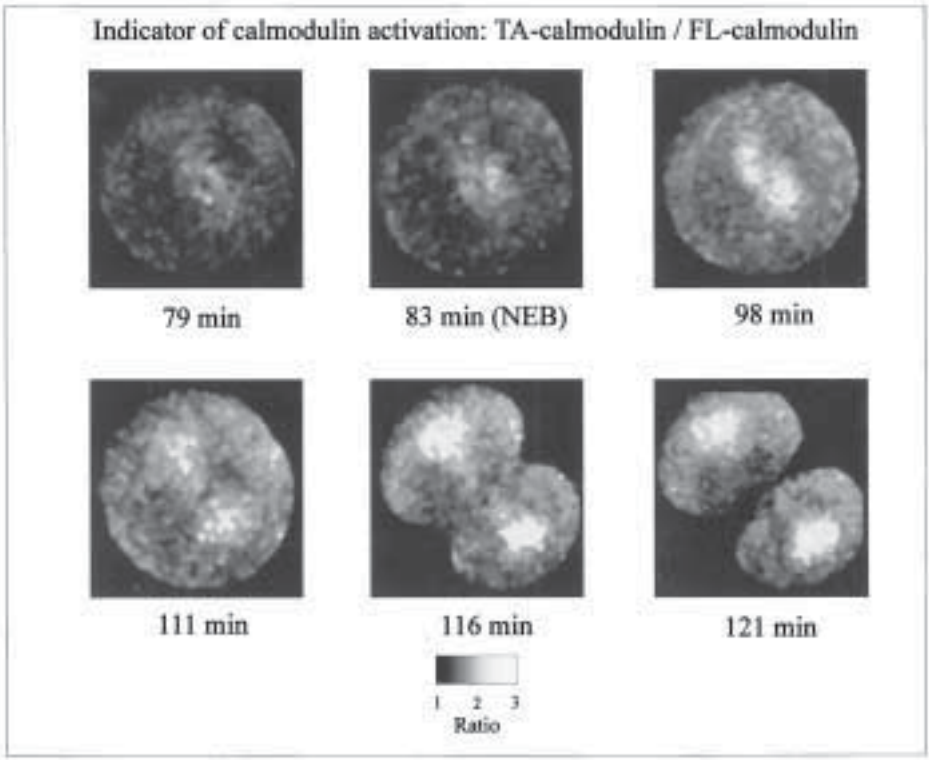


Fig. 17. Indicator of calmodulin activation using TA-calmodulin and FL-calmodulin. Simultaneous use of the activation-sensitive and activation-insensitive derivative allows us to distinguish between activation and local concentration changes of calmodulin. The sequence shows mitotic events from 79–121 min.

FL-calmodulin is an inert fluorophore. If the [$Ca^{2+}$] is changed from 10 n*M* – 0.1 m*M* at physiological ionic strength and pH, the fluorescence intensity of FL-calmodulin changes by less than 5%. No further change occurs on peptide target binding. Similar observations were made with Cy5-calmodulin and Texas Red-calmodulin. Thus FL-, Cy5-, and Texas Red-calmodulins report calmodulin localization in the cell. In contrast, if calmodulin is labeled at $Lys_{75}$ with the environmentally sensitive TA-Cl probe ***(11)***, $Ca^{2+}$-binding results in a 5.5-fold increase of fluorescence intensity and target binding may cause a further twofold increase ***(12)***. TA-calmodulin fluorescence thus reports the interactions of calmodulin, as well as its concentration changes in the cell. In order to distinguish between the two events, TA- and FL-calmodulin can both be applied to the same cell and their fluorescence emissions are scanned simultaneously.

5. It is instructive to compare the target binding and enzyme activation properties of fluorescently labeled calmodulin with unlabeled calmodulin. TA-calmodulin and

the significantly less-bright $Lys_{75}$-labeled DANSYL-calmodulin bind to targets with an approx threefold increased dissociation constant and act as an activator of cyclic-AMP phosphodiesterase similar to unmodified calmodulin. $Lys_{75}$-modified calmodulins appear to act as a competitive inhibitor of smooth muscle myosin light-chain kinase ***(13)***. They do, however, activate calmodulin-dependent protein kinase II auto- and substrate phosphorylation (Török, K. and Fraser, C., unpublished data). It is thus expected that $Lys_{75}$-labeled calmodulins are accurate reporters of calmodulin movements and activities in the cell. The inhibitory property can either be taken advantage of or countered by trace-level application of the fluorescent calmodulin in the cell.

## References


1. Cohen, P. and Klee, C. B., eds. (1988) *Calmodulin*. Elsevier, New York.
2. Mann, D. and Vanaman, T. C. (1987) Specific chemical modification as a probe of calmodulin function. *Methods Enzymol.* **139,** 417–433.
3. Zhang, M. and Vogel, H. J. (1993) NMR studies of the pKa's of the lysine sidechains in calmodulin. *J. Biol. Chem.* **268,** 22,420–22,428.
4. Török, K., Lane, A. N., Martin, S. R., Janot, J.-M., and Bayley, P. M. (1992) Effects of calcium binding on the internal dynamic properties of bovine brain calmodulin, studied by NMR and optical spectroscopy. *Biochemistry* **31,** 3452–3462.
5. Giedroc, D. P., Puett, D., Sinha, S. K., and Brew, K. (1987) Calcium effects on calmodulin lysine reactivities. *Arch. Biochem. Biophys.* **252,** 136–144.
6. Selsted, M. E. (1997) HPLC methods for purification of antimicrobial peptides. *Methods Mol. Biol.* **78,** 17–33.
7. Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., and Udseth, H. R. (1990) New developments in biochemical mass spectrometry: electrospray ionisation. *Anal. Chem.* **62,** 882–899.
8. Mann, M. and Wilm, M. (1995) Electrospray mass spectrometry for protein characterization. *Trends Biochem. Sci.* **20,** 219–224.
9. Allen, G. (1989) Sequencing of proteins and peptides. Laboratory techniques in *Biochemistry and Molecular Biology* (Burdon, R. H. and van Knippenberg, eds.), Elsevier, Amsterdam.
10. Yost, R. A. and Boyd, R. K. (1990) Tandem mass spectrometry: quadrupole and hybrid instruments. *Methods Enzymol.* **193,** 154–200.
11. Cowley, D. J., O'Kane, E., and Todd, R. S. J. (1991) Triazinylaniline derivatives as fluorescence probes. Part 1. Absorption and fluorescence in organic solvents and in aqueous media in relation to twisted intramolecular charge-transfer state formation, H bonding and protic equilibria. *J. Chem. Soc. Perkin. Trans.* **2,** 1495–1500.
12. Török, K. and Trentham, D. R. (1994) Mechanism of 2-chloro-(ε-amino-$Lys_{75}$)-(6-(4-N,N-diethylamino-phenyl)-1,3,5-triazin-4-yl)-calmodulin interactions with smooth muscle myosin light chain kinase and derived peptides. *Biochemistry* **33,** 12,807–12,820.
13. Török, K., Cowley, D. J., Brandmeier, B. D., Howell, S., Aitken A., and Trentham. D. R. (1998) Inhibition of calmodulin-activated smooth muscle myosin light chain

kinase by calmodulin binding peptides and fluorescent (phosphodiesterase-activating) calmodulin derivatives. *Biochemistry* **37,** 6188–6198.

14. Teruel, M. N. and Meyer, T. (1997) Electroporation-induced formation of individual calcium entry sites in the cell body and processes of adherent cells. *Biophys. J.* **73,** 1785–1796.
15. Wilding, M., Török K., and Whitaker M. J. (1995) Activation-dependent and activation-independent localisation of calmodulin to the mitotic apparatus during the first cell cycle of the Lytenichus pictus embryo. *Zygote* **3,** 219–224.
16. Pawley, J., ed. (1989) *The Handbook of Biological Confocal Microscopy*. IMR Press, Madison, Wisconsin.
17. Török, K., Wilding, M., Groigno, L., Patel, R. D., and Whitaker, M. J. (1998) Spatial dynamics of calmodulin activation during mitosis in early sea urchin embryos. *Curr. Biol.* **8,** 692–699.

# Index

METHODS IN MOLECULAR BIOLOGY™ • 173
Series Editor: *John M. Walker*

# Calcium-Binding Protein Protocols

## *Volume II: Methods and Techniques*

Edited by

***Hans J. Vogel***

*Department of Biological Sciences, University of Calgary, Calgary, AB, Canada*

Calcium-binding proteins play an important role in a variety of vital biological processes, ranging from blood clotting and signal transduction in cells, to attaching proteins to membranes and serving as an integral source of calcium. In *Calcium-Binding Protocols—Volume 1: Reviews and Case Studies and Volume 2: Methods and Techniques*—Hans Vogel and a panel of leading researchers review the protein chemistry and behavior of this significant protein class, and provide a comprehensive collection of proven experimental techniques for their study both in vitro and in vivo. This second volume focuses on cutting-edge experimental techniques for studying the solution structure, stability, dynamics, calcium-binding properties, and biological activity of calcium-binding protein in general. In addition to enzymatic assays and more routine spectroscopic and protein chemistry techniques, there are also NMR approaches, thermodynamic analyses, kinetic measurements such as surface plasmon resonance, strategies for amino acid sequence alignments, and fluorescence methods to study the distribution of calcium and calcium-binding proteins in cells. The first companion volume, *Reviews and Case Histories* sets the stage for this volume by introducing the various classes of intra- and extracellular calcium-binding proteins and their mode of action.

Comprehensive and highly practical, the two volumes of *Calcium-Binding Protocols* provide experimental and clinical biologists with a host of advanced experimental methods that can be applied successfully to the study of both existing and newly discovered members of this critically important class of proteins.

## FEATURES

- **All major biophysical and protein methods to study calcium-binding proteins**
- **Detailed discussion of calcium-binding proteins in vitro and in vivo**
- **Methods using fluorescence spectroscopy, NMR, thermodynamic analysis, and kinetic measurements**
- **Many methods also applicable to proteins that do not bind to calcium**

## CONTENTS



ISBN 0-89603-689-8

9 780896 036895 90000

**Methods in Molecular Biology™ • 173**
**CALCIUM-BINDING PROTEIN PROTOCOLS**
***VOLUME II: METHODS AND TECHNIQUES***
**ISBN: 0-89603-689-8**
**humanapress.com**