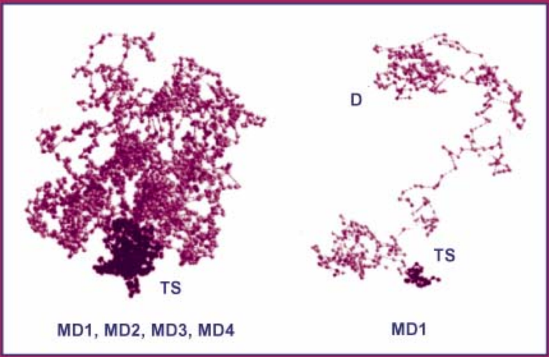
Methods in Molecular Biology™

VOLUME 168

Protein Structure, Stability, and Folding

Edited by

Kenneth P. Murphy



HUMANA PRESS

Protein Structure, Stability, and Folding

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, *Series Editor*

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, 2001*
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*
148. **DNA–Protein Interactions:** *Principles and Protocols* **(2nd ed.)**, edited by *Tom Moss, 2001*
147. **Affinity Chromatography:** *Methods and Protocols*, edited by *Pascal Bailon, George K. Ehrlich, Wen-Jian Fung, and Wolfgang Berthold, 2000*
146. **Mass Spectrometry of Proteins and Peptides,** edited by *John R. Chapman, 2000*
145. **Bacterial Toxins:** ***Methods and Protocols***, edited by *Otto Holst, 2000*
144. **Calpain Methods and Protocols**, edited by *John S. Elce, 2000*
143. **Protein Structure Prediction:** *Methods and Protocols*, edited by *David Webster, 2000*
142. **Transforming Growth Factor-Beta Protocols**, edited by *Philip H. Howe, 2000*
141. **Plant Hormone Protocols**, edited by *Gregory A. Tucker and Jeremy A. Roberts, 2000*
140. **Chaperonin Protocols**, edited by *Christine Schneider, 2000*
139. **Extracellular Matrix Protocols**, edited by *Charles Streuli and Michael Grant, 2000*
138. **Chemokine Protocols**, edited by *Amanda E. I. Proudfoot, Timothy N. C. Wells, and Christine Power, 2000*
137. **Developmental Biology Protocols, Volume III**, edited by *Rocky S. Tuan and Cecilia W. Lo, 2000*
136. **Developmental Biology Protocols, Volume II,** edited by *Rocky S. Tuan and Cecilia W. Lo, 2000*
135. **Developmental Biology Protocols, Volume I,** edited by *Rocky S. Tuan and Cecilia W. Lo, 2000*
134. **T Cell Protocols:** *Development and Activation*, edited by *Kelly P. Kearse, 2000*
133. **Gene Targeting Protocols,** edited by *Eric B. Kmiec, 2000*
132. **Bioinformatics Methods and Protocols,** edited by *Stephen Misener and Stephen A. Krawetz, 2000*
131. **Flavoprotein Protocols,** edited by *S. K. Chapman and G. A. Reid, 1999*
130. **Transcription Factor Protocols,** edited by *Martin J. Tymms, 2000*
129. **Integrin Protocols,** edited by *Anthony Howlett, 1999*
128. **NMDA Protocols,** edited by *Min Li, 1999*
127. **Molecular Methods in Developmental Biology:** Xenopus *and Zebrafish*, edited by *Matthew Guille, 1999*
126. **Adrenergic Receptor Protocols,** edited by *Curtis A. Machida, 2000*
125. **Glycoprotein Methods and Protocols:** *The Mucins*, edited by *Anthony P. Corfield, 2000*
124. **Protein Kinase Protocols,** edited by *Alastair D. Reith, 2001*
123. ***In Situ* Hybridization Protocols (2nd ed.),** edited by *Ian A. Darby, 2000*
122. **Confocal Microscopy Methods and Protocols,** edited by *Stephen W. Paddock, 1999*
121. **Natural Killer Cell Protocols:** *Cellular and Molecular Methods,* edited by *Kerry S. Campbell and Marco Colonna, 2000*
120. **Eicosanoid Protocols,** edited by *Elias A. Lianos, 1999*
119. **Chromatin Protocols,** edited by *Peter B. Becker, 1999*
118. **RNA–Protein Interaction Protocols,** edited by *Susan R. Haynes, 1999*
117. **Electron Microscopy Methods and Protocols,** edited by *M. A. Nasser Hajibagheri, 1999*
116. **Protein Lipidation Protocols,** edited by *Michael H. Gelb, 1999*
115. **Immunocytochemical Methods and Protocols (2nd ed.),** edited by *Lorette C. Javois, 1999*

METHODS IN MOLECULAR BIOLOGY™

Protein Structure, Stability, and Folding

Edited by

Kenneth P. Murphy

University of Iowa College of Medicine, Iowa City, IA

Humana Press **Totowa, New Jersey**

© 2001 Humana Press Inc.
999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise without written permission from the Publisher. Methods in Molecular Biology™ is a trademark of The Humana Press Inc.

The content and opinions expressed in this book are the sole work of the authors and editors, who have warranted due diligence in the creation and issuance of their work. The publisher, editors, and authors are not responsible for errors or omissions or for any consequences arising from the information or opinions presented in this book and make no warranty, express or implied, with respect to its contents.

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

Production Editor: Kim Hoather-Potter.
Cover design by Patricia F. Cleary.
Cover illustration: From Fig. 6 in Chapter 10 “Molecular Dynamics Simulations of Protein Unfolding/Folding” by Valerie Daggett.

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

Photocopy Authorization Policy:
Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Humana Press Inc., provided that the base fee of US $10.00 per copy, plus US $00.25 per page, is paid directly to the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license from the CCC, a separate system of payment has been arranged and is acceptable to Humana Press Inc. The fee code for users of the Transactional Reporting Service is: [0-89603-682-0/01 $10.00 + $00.25].

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™.

Protein structure, stability, and folding / edited by Kenneth P. Murphy
p. cm. — (Methods in molecular biology; v. 168)
Includes bibliographical references and index.
ISBN 0-89603682-0 (alk. paper)
1. Protein folding. 2. Proteins—Conformation. I. Murphy, Kenneth P. II. Methods in molecular biology (Clifton, N.J.) ; . v. 168.

QP551 .P697693 2001
572'.633—dc21 00-044883

Preface

In 1995, Humana Press published a book edited by Dr. Bret A. Shirley entitled *Protein Stability and Folding: Theory and Practice*. This book detailed the use of many of the most important techniques for studying the stability and folding of proteins, but emphasized the Practice more than the Theory. The current volume, *Protein Structure, Stability, and Folding*, is a companion to the earlier work. In this volume more emphasis is placed on the Theory, although chapters on exciting new experimental approaches are included. Together, the two volumes cover a broad range of experimental and computational approaches to studying and understanding protein structure, stability, and folding.

Since the publication of Shirley's work, there have been many exciting advances in both theory and experiment. Advances in the power of theoretical approaches now make it feasible to calculate many aspects of protein stability and dynamics from knowledge of the structure. The present book, *Protein Structure, Stability, and Folding* includes chapters on these calculations, as well as approaches for calculating conformational entropy and performing molecular dynamics simulations of protein unfolding. New experimental approaches are also being used to look at stability and folding. These include the use of co-solvents, new applications of hydrogen exchange techniques, temperature-jump methods for looking at fast folding events, and new strategies for mutagenesis experiments.

I would like to thank all the authors who contributed to this book for their hard work and extraordinary patience. I would also like to thank my colleagues in the Department of Biochemistry and especially the members of my laboratory for their encouragement in this project. Finally I would like to thank my wife, Lee, and my children, Kevin, Katie, and Ellen, whose love and faith make it all worthwhile.

Kenneth P. Murphy

Contents

Contributors

D. Wayne Bolen • *Department of Human Biological Chemistry, University of Texas Medical Branch, Galveston, TX*

Trevor P. Creamer • *Kentucky Center for Structural Biology and Department of Biochemistry, University of Kentucky, Lexington, KY*

Valerie Daggett • *Department of Medicinal Chemistry, University of Washington, Seattle, WA*

S. Walter Englander • *Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA*

Ernesto Freire • *Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, MD*

Carl Frieden • *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO*

James Hofrichter • *Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD*

Vincent J. Hilser • *Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX*

Enoch S. Huang • *Pfizer Discovery Technology Center, Pfizer, Inc., Cambridge, MA*

Beatrice M. P. Huyghues-Despointes • *Department of Medical Biochemistry, Texas A&M University, College Station, TX*

Kenneth P. Murphy • *Department of Biochemistry, University of Iowa College of Medicine, Iowa City, IA*

C. Nick Pace • *Department of Medical Biochemistry, Texas A&M University, College Station, TX*

Jay W. Ponder • *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO*

Andrew D. Robertson • *Department of Biochemistry, University of Iowa College of Medicine, Iowa City, IA*

J. Martin Scholtz • *Department of Medical Biochemistry, Texas A&M University, College Station, TX*

T. Sivaraman • *Department of Biochemistry, University of Iowa College of Medicine, Iowa City, IA*

1

Stabilization of Protein Structure

Kenneth P. Murphy

1. Introduction

The complex and interrelated reactions which make up living processes are dependent upon the presence of proteins, not only as catalysts, but in their role as structural molecules, as storage and carrier molecules, and as molecular motors. All of these functions require that the nascent amino-acid chain correctly folds into the biologically active, three-dimensional structure of the native state. In pioneering studies, Anfinsen showed that all the necessary information for the nascent chain to fold into the native structure is contained in the sequence of amino acids ***(1)***. In order for proteins to fold spontaneously into their native state, the native state must be lower in Gibbs energy than the unfolded state. From this observation, Anfinsen proposed that the native and unfolded states are in equilibrium, and that the native state is the global minimum in the Gibbs energy that can be obtained without breaking covalent bonds. This assumption is known as the thermodynamic hypothesis, because it suggests that protein folding is under thermodynamic—rather than kinetic—control.

It is important to recognize that the thermodynamic hypothesis involves two key concepts: 1) that the native state is the global minimum in Gibbs energy; and 2) that the unfolded and native states are in equilibrium. The first concept cannot be proven experimentally,

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

although it is amenable to computational studies. The second concept is well-documented experimentally. If the native state is not the global Gibbs energy minimum, then the barrier between the native and global-minimum structure is large enough that a thermodynamic treatment of protein stability is perfectly justified.

If the native and unfolded states are in equilibrium, then no additional factors are required to fold the nascent chain, and it is unnecessary to have folding occur cotranslationally. In fact, the idea that proteins fold as the chain emerges from the ribosome cannot generally be true. It has been shown that the N-terminal domain of some proteins is unstable in the absence of the C-terminal domain. Thus the N-terminal domain—although synthesized first—does not fold independently and cannot fold cotranslationally. The molecules involved in protein folding, such as chaperones, must therefore serve the role of preventing off-pathway processes such as aggregation, rather than catalyzing folding per se.

The goal of this chapter is to provide an overview of the forces responsible for the lower Gibbs energy of the native state. An understanding of the forces responsible for the stabilization of protein structure is critical to an understanding of specificity in protein folding and binding. An understanding of these forces is also key to the prediction of protein structure, the structure-based design of pharmaceuticals, the design of proteins with enhanced stability for use in industrial and pharmaceutical processes, and rationalizing protein function.

2. Thermodynamic Description of Protein Folding

Because the native and unfolded states are in equilibrium, the folding reaction can be quantified in terms of thermodynamics. The contribution of various interactions to these thermodynamic quantities can also be explored as a means of understanding how they contribute to protein stability. The equilibrium between the native (N) and unfolded (U) states is defined by the equilibrium constant, K, as:

$$\mathrm{K} = \frac{[\mathrm{U}]}{[\mathrm{N}]} \tag{1}$$

The difference in free energy between the unfolded and native states is then:

$$\Delta G° = – R\ T \ln K \tag{2}$$

where *R* is the universal gas constant and *T* is the absolute temperature in Kelvin. Note that the equilibrium constant in **Eq. 1** is for the unfolding reaction, so that a positive ΔG° indicates that the native state is more stable than the unfolded state.

The free energy is composed of both enthalpic and entropic contributions:

$$\Delta G° = \Delta H° – T\ \Delta S° \tag{3}$$

where ΔH° is the enthalpy change and ΔS° is the entropy change upon unfolding. However, as with other reactions which take place in water, both the ΔH° and ΔS° are strongly temperature-dependent so that the free-energy change is better written as:

$$\Delta G° = \Delta H°_R – T\Delta S°_R + \Delta C_p \left\{ (T – T_R) – T\ ln(\frac{T}{T_R}) \right\} \tag{4}$$

where the subscript *R* indicates the value of ΔH° and ΔS° at a reference temperature, T_R, and ΔC_p is the heat-capacity change.

The result of the significant ΔC_p is that the free energy is not a linear function of temperature. Rather, it shows significant downward curvature and has a maximum value at some temperature, often near physiological. The temperature of maximum stability occurs where the ΔS° is equal to zero. The curvature is sometimes significant enough that the protein can unfold upon cooling, a phenomenon known as cold denaturation ***(2–4)***. These features are illustrated in **Fig. 1**, which shows the ΔG° vs temperature for a protein of 150 amino-acid residues using the average values of ΔH°, ΔS°, and ΔC_p per residue for globular proteins taken from a recent review ***(5)***.

In this chapter, the contributions of the dominant interactions to the ΔG° as a function of temperature are reviewed.

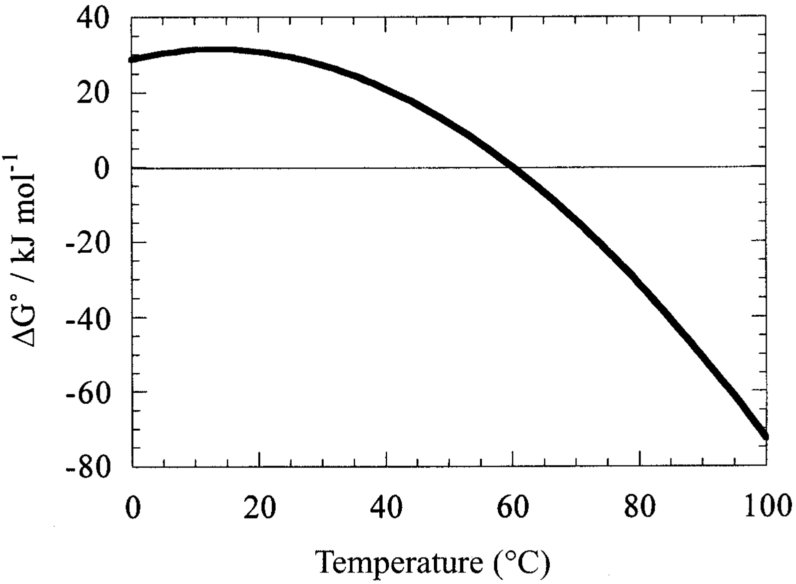


Fig. 1. ΔG° as a function of temperature for an average protein of 150 amino-acid residues with a T_m = 60°C, ΔH° = 438 kJ mol^{-1}, and ΔC_p = 8.7 kJ K^{-1} mol^{-1}. Note that the ΔG° goes through a maximum near 15°C and then decreases upon further decrease of temperature.

3. Description of the Dominant Forces in Protein Folding

The nature of the contribution of various interactions to protein folding and stability has been of interest for some time. In 1959 Walter Kauzmann wrote a seminal review in which he gave a physical description of these interactions and discussed the probable contribution of each ***(6)***. More recently, Ken Dill reviewed the data since Kauzmann's review ***(7)***, and concluded that Kauzmann's analysis has largely stood the test of time.

Typically the interactions which are considered to dominate protein stability are the hydrophobic effect, hydrogen bonding, and configurational entropy. Before discussing the contribution of these interactions to protein stability, it is helpful to define precisely what is meant by each term.

3.1. The Hydrophobic Effect

The hydrophobic effect generally refers to the low solubility of hydrophobic (i.e., apolar) compounds in water. In order to quantitate this effect, it is necessary to define not only the final state—i.e., hydrocarbon dissolved in water—but also an initial state. Hydrophobicity scales are usually based on a liquid hydrocarbon or liquid alcohol initial state so that the free energy for the hydrophobic effect would be defined as the $\Delta G°$ of transfer from the liquid hydrocarbon to aqueous solution. While the liquid initial state has been suggested as the best definition of the hydrophobic effect ***(8)***, it is difficult to apply directly to protein stability because the protein interior is not liquid-like ***(9,10)***. Thus, one would need to include the energetics of transfer from the protein interior to the liquid in order to fully account for the contribution of hydrophobic groups to protein stability.

For the purposes of this chapter, the hydrophobic effect will be defined as the energetics associated with the transfer of hydrophobic surface from the protein interior to water. This includes the energetics associated with changes in van der Waals interactions experienced by the hydrophobic groups, as well as the energetics associated with the restructuring of water which take place upon hydrating hydrophobic groups. Using this definition, it is difficult to make direct measurements of the hydrophobic contribution to protein stability. Yet, as discussed in **Subheading 4.1.**, this contribution can be estimated from the energetics of transfer of hydrophobic groups from amino-acid crystals to water as well as from studies of the energetics of protein stabilization themselves.

3.2. Hydrogen Bonding

As with the hydrophobic effect, hydrogen bonding can be defined in various ways. This definition will determine what is meant by the contribution of “hydrogen bonding” to protein stability. In general, hydrogen bonding refers to the partial sharing of a hydrogen atom between a donor atom to which it is covalently bound, and an acceptor atom which has a lone pair of electrons. In proteins, the intramolecular hydrogen bonding competes with hydrogen bonding to water molecules. It is this competition that determines the contribution of hydrogen bonding to protein stability.

For the purposes of this chapter, the contribution of hydrogen bonding will be defined in terms of this competition—i.e., it is the difference between the energetics of hydrogen bonding groups interacting with each other within the folded protein and interacting with solvent water in the unfolded protein. It is important to note that within this definition the contribution of van der Waals interactions between the protein interior and the hydrogen-bonded pair also contributes to hydrogen bonding.

Other definitions of the contribution of hydrogen bonding are possible, and have been used. For example, in a mutagenesis experiment, a hydrogen-bonded side chain can be replaced by a side chain which cannot form hydrogen bonds (e.g., replacing tyrosine with phenylalanine). The difference in the folding energetics between these two proteins can then be defined as the contribution of a hydrogen bond ***(11–13)***.

3.3. Configurational Entropy

In an unfolded protein, the polypeptide chain can adopt different rotomeric positions around ϕ and ψ torsional angles, and side chains can adopt different rotomers around χ angles. Upon adopting the native state, nearly all ϕ and ψ angles are restricted to a single position, as are the majority of χ angles. This loss of degrees of freedom translates into a loss of configurational entropy. It is the loss of configurational entropy which must be overcome by favorable interactions in order to fold a stable protein.

3.4. Other Contributions

Protein stability is affected through environmental factors such as temperature, pH, and salt. Temperature affects stability by changing the contributions of the hydrophobic effect, hydrogen bonding, and configurational entropy. In contrast, the effects of factors such as pH and salt arise from linked equilibria ***(14)***—protons or ions bind to the native and denatured states with different affinities. A complete understanding of protein stability also requires knowledge of these linkage effects.

4. Evaluating Contributions

The contribution of different interactions to the stability of globular proteins has been evaluated in a number of different ways, both experimentally and computationally. Computational approaches are particularly well-suited for evaluating configurational entropy, as discussed in **Subheading 4.3.**

Experimental studies can generally be divided into two types: model compound and mutational studies. In model compound studies the goal is to study a system which captures most of the relevant features of protein stability, but is simple enough to be more readily interpreted. The various types of model compound studies have been described previously ***(15–17)***, and are not reviewed in this chapter. However, the results of such studies will be discussed here.

4.1. The Hydrophobic Effect

Model compound studies of the transfer of hydrophobic compounds into water have resulted in a thorough understanding of the contribution of the hydrophobic effect to protein stability. Whether the transfer into water is from the gas, liquid, or solid phase, it is evident that the hydrophobic effect is accompanied by a significant increase in heat capacity and, at 25°C, a significant decrease in entropy (for a review, *see* **ref. *18***). This temperature-dependent decrease in entropy approaches zero near 112°C, assuming a temperature-independent ΔC_p ***(19,20)***. The large ΔC_p results in a significant curvature to $\Delta G°$ as a function of temperature.

In contrast to ΔC_p and $\Delta S°$, the value of the $\Delta H°$ associated with the hydrophobic effect depends on whether the transfer is from the gaseous, liquid, or aqueous phase ***(21)***, as well as the type of compound being transferred. With liquid hydrocarbons, the $\Delta H°$ of the

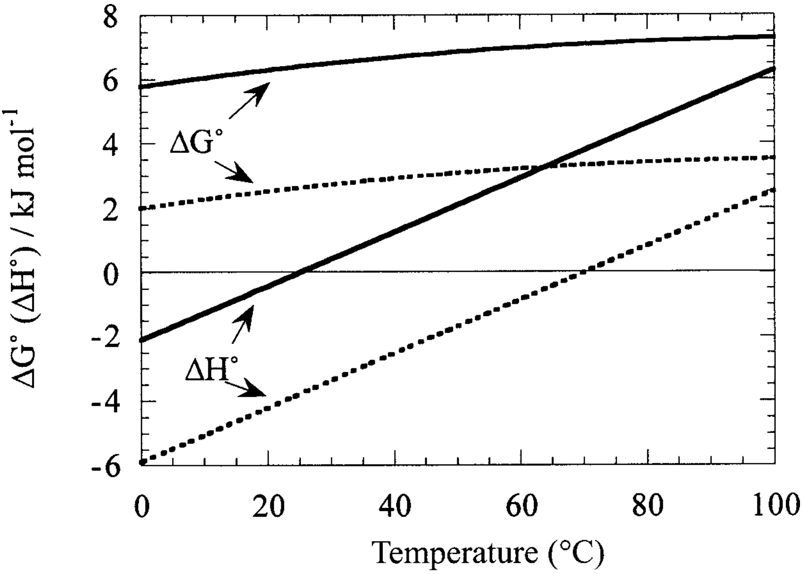


Fig. 2. The contribution of a methyl group to the ΔG° and ΔH° of the hydrophobic effect for liquid hydrocarbons (solid lines) and amino-acid crystals (dashed lines). Note that the ΔS° for both processes is equal (not shown), so that the decreased ΔG° for the crystals results from the decreased ΔH°.

hydrophobic effect at 25°C is near zero ***(22)***, but with crystalline, cyclic dipeptides it is negative ***(23)***. The consequence of this negative ΔH° is that the ΔG° for the hydrophobic effect as it applies to cyclic dipeptides is less than what is estimated from the dissolution of hydrocarbon liquids (*see* **Fig. 2**). This negative ΔH° for the hydrophobic effect also seems to apply to proteins ***(5,21,24,25)***, suggesting that the hydrophobic effect contributes less to the stability of globular proteins than originally believed. This finding is supported by protein stability studies ***(5,21,26,27)***.

Mutational studies which attempt to examine the hydrophobic effect show a wide range of values. This is presumably the result of packing differences between side chains (*see* **ref. *28***). The results of mutational studies mostly illustrate the difficulty of interpreting small changes in complex systems such as proteins. This is especially true when only the effects on the melting temperature, or ΔG°, is studied. The complexity of the interactions is revealed more clearly in studies that examine the effects of mutations on ΔH° and ΔS° in addition to ΔG° ***(29)***.

4.2. Hydrogen Bonding

As with the hydrophobic effect, both model compound and protein mutant studies have provided important information on the contribution of hydrogen bonding to protein stability. It might seem at first that hydrogen bonding would make little contribution to protein stability, since intramolecular hydrogen bonds broken upon protein unfolding are replaced by hydrogen bonds with water. However, hydrogen-bond strength depends not only on the presence of donor and acceptor groups, but on the chemical nature of the groups involved, the local environment in which the hydrogen bond is present, and the relative concentration of the groups.

One approach to ascertaining the contribution of hydrogen bonding to protein stability has been to study the interactions in solution of solutes that can form hydrogen bonds. Such compounds include urea, carboxylic acids, alcohols, and amides. These studies suffer from lack of information on the nature of the interaction between solutes. For example, interactions between alcohols can arise from hydrogen bonding between the hydroxyl groups or from a hydrophobic association between the aliphatic tails.

A classic illustration of the problem of interpreting solution interaction studies comes from the work of Klotz and Franzen ***(30)***, who studied the dimerization of N-methyl-acetamide as a model for hydrogen bonding in proteins. Observing no temperature dependence to the association constant, they concluded that the ΔH° of hydrogen-bond formation between amide groups was zero. This conclusion was quickly called into question by Némethy and Scheraga ***(31)***, who pointed out that the ΔH° from the hydrophobic interaction between the methyl groups could cancel the ΔH° of hydrogen bonding. While this suggestion was supported by subsequent solution studies ***(32–34)***, the idea that hydrogen bonding in proteins makes no enthalpic contribution to protein stabilization persists (*see* **ref.** ***35***).

Subsequent studies on the solution interaction of δ-valerolactam ***(36)*** and diketopiperazine ***(34,37)*** supported the conclusion of Némethy and Scheraga. Both of these compounds have *cis* peptide bonds, so that the hydrophobic groups would be directed away from each other in a hydrogen-bond interaction. The ΔH° of interaction in these compounds is significant and negative, suggesting that hydrogen bonds between peptide groups are enthalpically favorable in water. Nevertheless, the lack of structural data on the solution interactions limits the interpretation.

The energetics of transfer of amino-acid compounds from a crystalline solid into aqueous solution are more readily interpreted, because the interactions in the crystal can be determined from X-ray crystallography, while the solutes interact only with solvent water in dilute solution. Studies of the dissolution of a homologous series of cyclic dipeptides—in which cyclization removes the zwiterrionic charges—also indicate that hydrogen bonding of peptide groups within proteins is enthalpically stabilizing relative to the peptide groups interacting with solvent ***(21)***. Further studies of cyclic dipeptides containing serine side chains indicate that hydrogen bonding between the side chain hydroxyl and the backbone is also enthalpically stabilizing, but only about one-half as strong as the peptide-peptide interaction ***(38)***.

The enthalpic stabilization resulting from side chain to backbone hydrogen bonds is supported by studies of alanine to serine mutations in T4 lysozyme which show increased enthalpies of unfolding, although the melting temperatures decrease ***(39)***. These studies indicate that there is an entropic penalty of forming hydrogen bonds between side chains. Recent studies of an engineered salt bridge in a rubridoxin variant yield similar results ***(40)***. These studies revealed that a surface salt-bridge between side chains had little or no effect on the stability of the protein, but a salt bridge between a side chain and the backbone was stabilizing by 6.3 kJ mol^{-1}. Mutational studies in RNase T1 also indicate that hydrogen bonds stabilize protein structure ***(13)***.

4.3. Configurational Entropy

Configurational entropy, the primary destabilizing contribution to protein stability, arises from the additional rotomeric states accessible to amino-acid side chains and the protein backbone in the unfolded state. It is also important to realize that the configurational entropy of the backbone is dependent on the side chain, as is evident from Ramachandran plots ***(41,42)***.

While experimental approaches can be used to determine the difference in backbone entropy for different amino-acid residues ***(43)***, computational methods are the best approach for estimating the contribution of configurational entropy to protein stability. These computational approaches include statistical surveys of rotomer populations in proteins of known structure ***(44)***, Monte Carlo simulations ***(42,45)***, and exhaustive calculations ***(46,47)***. These studies indicate that the backbone contributes between 9 J K^{-1} mol^{-1} per residue (for Val and Ile) and 27 J K^{-1} mol^{-1} per residue (for Gly) to the configurational entropy ***(47)***, while the side chains contribute between 0 and 30 J K^{-1} mol^{-1} ***(46)***. Thus, the configurational entropy represents a considerable challenge to the folding of globular proteins.

5. Conclusion

The stabilizing contributions that arise from the hydrophobic effect and hydrogen bonding are largely offset by the destabilizing configurational entropy. The hydrophobic effect is strongly temperature-dependent, and is considerably weaker and perhaps even destabilizing at low temperatures than at elevated temperatures. The contribution of various interactions for a “typical” protein is illustrated in **Fig. 3**.

Because the stability of a protein is a balancing act between large, opposing contributions, it is easily tipped in one direction or the

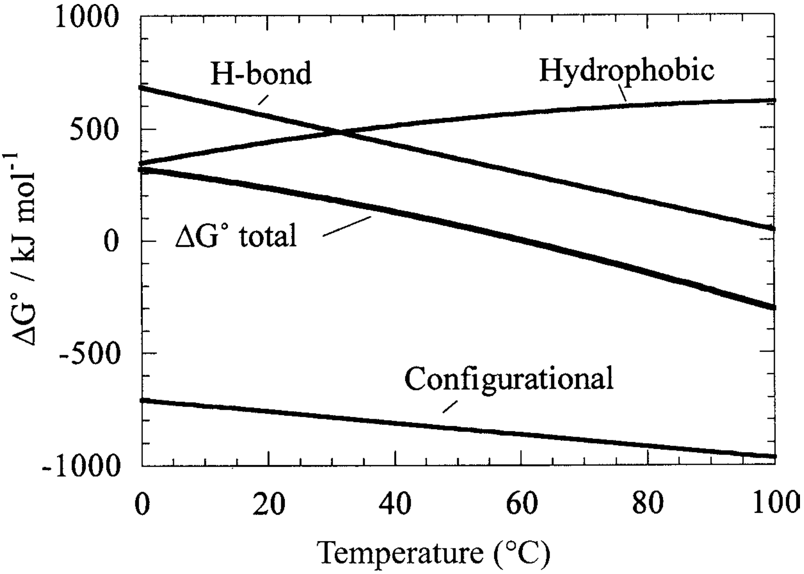


Fig. 3. Representative contributions of interactions to the overall stabilization of a globular protein. Both the hydrophobic effect and hydrogen bonding make large, favorable contributions to the overall ΔG°, while the configurational entropy makes a large, unfavorable contribution. The result is the modest stability characteristic of globular proteins.

other by relatively small changes in the environment. These can be changes in the temperature, pH, or ligand concentration, or the addition of denaturants or stabilizing cosolvents. The ability to switch from stabilizing to destabilizing conditions relatively easily conveniently allows the study of the folding reaction.

While the exact contribution of different interactions to the stability of globular proteins remains in question, our understanding seems to be refined enough to allow for the reasonable prediction of the overall folding thermodynamics. What remains is to refine this understanding into precise predictions of melting temperatures and their dependence on small environmental changes.

References

1. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181,** 223–230.

2. Brandts, J. F. (1964) The thermodynamics of protein denaturation. I. The denaturation of chymotrypsinogen. *J. Am. Chem. Soc.* **86,** 4291–4301.
3. Lumry, R., Biltonen, R., and Brandts, J. F. (1966) Validity of the “two-state” hypothesis for conformational transitions of proteins. *Biopolymers* **4,** 917–944.
4. Privalov, P. L. (1990) Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **25,** 281–305.
5. Robertson, A. D. and Murphy, K. P. (1997) Protein Structure and the energetics of protein stability. *Chem. Rev.* **97,** 1251–1267.
6. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14,** 1–63.
7. Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry* **29,** 7133–7155.
8. Herzfeld, J. (1991) Understanding hydrophobic behavior. *Science* **253,** 88.
9. Chothia, C. (1984) Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **53,** 537–572.
10. Harpaz, Y., Gerstein, M., and Chothia, C. (1994) Volume changes on protein folding. *Structure* **2,** 641–649.
11. Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P., and Matthews, B. W. (1987) Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature* **330,** 41–46.
12. Shirley, B. A., Stanssens, P., Hahn, U., and Pace, C. N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31,** 725–732.
13. Myers, J. K. and Pace, C. N. (1996) Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **71,** 2033–2039.
14. Wyman, J. and Gill, S. J. (1990) Binding and Linkage: The Functional Chemistry of Biological Macromolecules, University Science Books, Mill Valley, CA.
15. Gill, S. J. (1988) Calorimetry of model compounds that mimic biological molecules. *J. Chem. Thermodynamics* **20,** 1361–1382.
16. Murphy, K. P. (1995) Noncovalent forces important to the conformational stability of protein structures, in *Protein Stability and Folding* (Shirley, B. A., ed.), Humana Press, Totowa, NJ, pp. 1–34.
17. Murphy, K. P. (1997) Model Compounds and the Interpretation of Protein-Ligand Interactions, in *Thermodynamics and Structure Based*

Drug Design (Ladbury, J. E. and Connelly, P. R., eds.), R. G. Landes Co., Austin, TX, pp. 85–109.

18. Privalov, P. L. and Gill, S. J. (1988) Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* **39,** 191–234.
19. Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. USA* **83,** 8069–8072.
20. Murphy, K. P., Gill, S. J., and Privalov, P. L. (1990) Thermodynamic features common to protein unfolding and dissolution of hydrophobic compounds in water. *Biophys. J.* **57,** 407a.
21. Murphy, K. P. and Gill, S. J. (1991) Solid model compounds and the thermodynamics of protein unfolding. *J. Mol. Biol.* **222,** 699–709.
22. Gill, S. J., Nichols, N. F., and Wadsö, I. (1976) Calorimetric determination of enthalpies of solution of slightly soluble liquids II. Enthalpy of solution of some hydrocarbons in water and their use in establishing the temperature dependence of their solubilities. *J. Chem. Thermodynamics* **8,** 445–452.
23. Murphy, K. P. and Gill, S. J. (1990) Group additivity thermodynamics for dissolution of solid cyclic dipeptides into water. *Thermochim. Acta* **172,** 11–20.
24. Murphy, K. P. and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* **43,** 313–361.
25. Xie, D. and Freire, E. (1994) Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. *Proteins* **19,** 291–301.
26. Murphy, K. P. (1994) Hydration and convergence temperatures: On the use and interpretation of correlation plots. *Biophys. Chem.* **51,** 311–326.
27. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J.* **10,** 75–83.
28. Lee, B. (1993) Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size. *Protein Sci.* **2,** 733–738.
29. Sturtevant, J. M. (1994) The thermodynamic effects of protein mutations. *Curr. Opin. Struct. Biol.* **4,** 69–78.
30. Klotz, I. M. and Franzen, J. S. (1962) Hydrogen bonds between model peptide groups in solution. *J. Am. Chem. Soc.* **84,** 3461–3466.

31. Némethy, G. and Scheraga, H. A. (1962) The Structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. *J. Phys. Chem.* **66,** 1773–1789.
32. Susi, H. and Ard, J. S. (1969) Hydrophobic interactions and hydrogen bonding of ε-caprolactam in aqueous solution. *J. Phys. Chem.* **73,** 2440–2441.
33. Suurkuusk, J. and Wadsö, I. (1974) Design and testing of an improved precise drop calorimeter for the measurement of the heat capacity of small samples. *J. Chem. Thermodynamics* **6,** 667–679.
34. Gill, S. J. and Noll, L. (1972) Calorimetric study of association of diketopiperazine in water. *J. Phys. Chem.* **76,** 3065–3068.
35. Honig, B. and Yang, A.-S. (1995) Free energy balance in protein folding. *Adv. Protein Chem.* **46,** 27–58.
36. Susi, H., Timasheff, S. N., and Ard, J. S. (1964) Near infrared investigation of interamide hydrogen bonding in aqueous solution. *J. Biol. Chem.* **239,** 3051–3054.
37. Murphy, K. P. and Gill, S. J. (1989) Calorimetric measurement of the enthalpy of dissolution of diketopiperazine in water as a function of temperature. *Thermochim. Acta* **139,** 279–290.
38. Habermann, S. M. and Murphy, K. P. (1996) Energetics of hydrogen bonding in proteins: a model compound study. *Protein Sci.* **5,** 1229–1239.
39. Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Heinz, D. W., and Matthews, B. W. (1993) Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala→Ser and Val→Thr substitutions in T4 lysozyme. *Biochemistry* **32,** 11,363–11,373.
40. Strop, P. and Mayo, S. L. (2000) Contribution of surface salt bridges to protein stability. *Biochemistry* **39,** 1251–1255.
41. Némethy, G., Leach, S. J., and Scheraga, H. A. (1966) The influence of amino acid side chains on the free energy of helix-coil transitions. *J. Phys. Chem.* **70,** 998–1004.
42. Creamer, T. P. and Rose, G. D. (1992) Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. USA* **89,** 5937–5941.
43. O'Neil, K. and DeGrado, W. (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250,** 646–651.

44. Pickett, S. D. and Sternberg, M. J. E. (1993) Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* **231,** 825–839.
45. Creamer, T. P. and Rose, G. D. (1994) α-helix-forming propensities in peptides and proteins. *Proteins* **19,** 85–97.
46. Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994) Entropy changes in biological processes: loss of side chain configurational entropy in binding and folding. *Proteins* **20,** 68–84.
47. D’Aquino, J. A., Gómez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25,** 143–156.

2

Protein Stabilization by Naturally Occurring Osmolytes

D. Wayne Bolen

1. Introduction

Natural selection is believed to be an unforgiving and relentless force in the evolution of life on earth. An organism that cannot adapt to a changing environment or an environment hostile to cell functions is at risk as a species. So it is important to understand the mechanisms used by plants, animals, and microorganisms in adapting to environments in the biosphere that would ordinarily denature proteins or otherwise cause disruption of life-giving cellular processes. These hostile environments involve such stresses as extremes of temperature, cellular dehydration, desiccation, high extracellular salt environments, and even the presence of denaturing concentrations of urea inside cells ***(1)***. It has been recognized for some time that many plants, animals, and microorganisms that have adapted to environmental extremes also accumulate significant intracellular concentrations of small organic molecules ***(1–4)***. From these (and other) observations comes the hypothesis that these small organic molecules, called osmolytes, have the ability to protect the cellular components against denaturing environmental stresses ***(1–5)***. In this chapter, we seek to understand the molecular-

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy

level phenomena involving proteins and the naturally occurring osmolytes that result in the stabilization of proteins against denaturation stresses.

The disaccharide trehalose is the principal osmolyte selected to protect the resurrection plant against desiccation, and it is the osmolyte selected by nature to protect tardigrades (primitive arthropods) against the deleterious effects of desiccation ***(6,7)***. This is only one example in which organisms from different kingdoms have been subjected to the same selective pressure (desiccation) and have settled on the same osmolyte for protection against that stress. Such cases have been cited as examples of convergent evolution ***(1)***. If convergent evolution indeed selects for particular organic compounds as cell-component protectants, then the property or properties of the osmolyte that bestow the protection must be highly valued evolutionarily. Identification of an evolutionarily selected property or properties of osmolytes that protect proteins against denaturing stresses should provide valuable insight into the fundamental problem of protein stabilization, a problem that plagues the pharmaceutical industry, biotechnology, and everyday research.

Satisfying biological needs and biological constraints are a necessity in adaptation, and it is most enlightening to approach the issue of protein stabilization in that context. In the natural selection of a mechanism that will enable an organism to adapt to denaturing stresses, two conditions must be met: (1) the mechanism must provide for the stabilization of cellular proteins and other cell components against the denaturing environmental stress; and (2) the mechanism must provide the stabilization without significantly affecting the functional activity of the proteins and other cellular components ***(1,8,9)***. This places stringent constraints on the types of mechanisms appropriated through natural selection. The condition of not affecting the functional activity of proteins and other cellular components is just as important as providing for stabilization, for there is no selective advantage for a mechanism that solves the stabilization problem while creating problems with cellular function. Our goal in this chapter is to investigate how physical and/or chemical properties of osmolytes may be responsible for the stabilization of proteins, and to relate these properties to the biology of adaptation.

2. Naturally Occurring (Protecting and Nonprotecting) Osmolytes

Intracellular accumulation of organic solutes creates an osmotic pressure that can affect cell volume, and this osmotic effect is the reason these solutes are often referred to as organic osmolytes ***(10)***. Small organic osmolytes were selected by nature to protect the cell and cellular components against particular denaturing stresses; however, not all osmolytes protect cell components. An important nonprotecting osmolyte is urea, which accumulates in mammalian renal medulla cells and also in the cells of sharks, rays, and the coelacanth ***(1)***. The intracellular proteins in urea-enriched cells are just as susceptible to the strong denaturing properties of urea as the proteins in cells that do not accumulate urea. What permits urea-rich cells to thrive is the intracellular presence of methylamine osmolytes—compounds that have the ability to offset (i.e., protect against) the deleterious effects of urea ***(8,11,12)***. Urea is an osmolyte in a class by itself—it is a nonprotecting osmolyte.

In contrast to urea, the array of naturally occurring organic osmolytes that are classified as protecting osmolytes fall into three general chemical classes: (1) the polyols, which include glycerol, sucrose, trehalose, and certain other sugars; (2) certain amino acids, including proline and glycine; and (3) particular methylamines, such as sarcosine, trimethylamine N-oxide (TMAO), and betaine. A more extensive listing of protecting osmolytes can be found in ***(1)***. Assuming that environmental stresses are the sources of selective pressure for the accumulation of organic osmolytes in an organism, particular classes of osmolytes appear to have been selected for their ability to oppose particular stress conditions. For example, polyols appear to be particularly good at protecting organisms against stresses such as temperature extremes and dehydration or desiccation, and amino acids appear to have been selected to protect cells against extracellular environments that are high in salt concentration ***(1)***. By contrast, the methylamine osmolytes seem to have the ability to protect the cellular components of urea-rich cells against the deleterious effects of urea ***(1)***.

A general conclusion that might be drawn from these observations is that protecting osmolytes are stress-specific, and one should not expect an osmolyte that is good at protecting against one denaturing stress to be particularly effective in protecting against a different denaturing stress. Yet it would be wrong to conclude that a different basic mechanism for protection is operative with each protecting osmolyte or osmolyte class. There is considerable similarity in the mechanism by which the organic osmolytes protect against denaturing stresses—a commonality that reveals itself in the property of preferential exclusion ***(13–24)***.

3. Essential Physico-Chemical Observations

3.1. Preferential Exclusion: A Common Feature of Protecting Osmolytes

Sucrose has long been used in biochemistry to stabilize proteins, and in an effort to understand this process, Lee and Timasheff ***(25)*** evaluated the binding of sucrose to proteins by equilibrium dialysis. Favorable binding of a ligand to protein results in the ligand concentration in the protein-containing compartment that is the sum of the concentrations of unbound ligand and ligand bound to protein, while in the compartment lacking protein only unbound ligand concentration is possible. At equilibrium, the chemical potentials of unbound ligand in both compartments are identical. Thus, favorable ligand binding results in a greater concentration of ligand in the protein-containing compartment than in the compartment lacking protein.

In their equilibrium dialysis experiment, Lee and Timasheff found that as a ligand, sucrose is observed to have a greater concentration in the compartment lacking protein than it has in the compartment containing protein—i.e., the results are the exact opposite of those expected for favorable ligand binding to protein ***(25)***. This result unequivocally shows that sucrose does not bind to proteins, and it is preferentially excluded from the vicinity of the protein. This preferential exclusion phenomenon has been found for the naturally occurring protecting osmolytes and is a common defining feature of these solutes.

Obviously, preferential exclusion of a ligand from the immediate vicinity of a protein implies that the concentration of ligand in the immediate volume element surrounding the protein is lower than its concentration in the bulk phase. Clearly, if ligand concentration is lower in this volume element than in the bulk phase, then water concentration in this volume element necessarily must be higher than it is in the bulk phase. Thus, preferential exclusion of ligand in the volume element is equivalent to saying the protein is preferentially hydrated. Thus, the terms *preferential exclusion* of ligand and *preferential hydration* of the protein refer to the same phenomenon, and are often used interchangeably.

3.2. Gibbs Energy Implications of Preferential Exclusion

By definition, favorable ligand binding to a protein always results in a thermodynamic decrease in the Gibbs energy for the process—i.e., the standard chemical potential for the interacting protein-ligand complex is lower than the sum of the individual standard chemical potentials of free ligand and free protein (ΔG° is negative). By contrast, preferential exclusion causes an increase in ΔG°, and the standard chemical potential of the sucrose-protein system is greater than the sum of the standard chemical potentials of free ligand and free protein ***(26,27)***. Using transfer Gibbs energy measurements, we have shown that the thermodynamic effect of stabilizing osmolytes on the equilibrium between native and unfolded protein gives a unique Gibbs energy diagram that illustrates how protein stabilization arises from preferential exclusion ***(27–29)***. Arakawa and Timasheff originally presented the basic features of the Gibbs energy diagram, reproduced here in **Fig. 1**, and based on preferential exclusion.

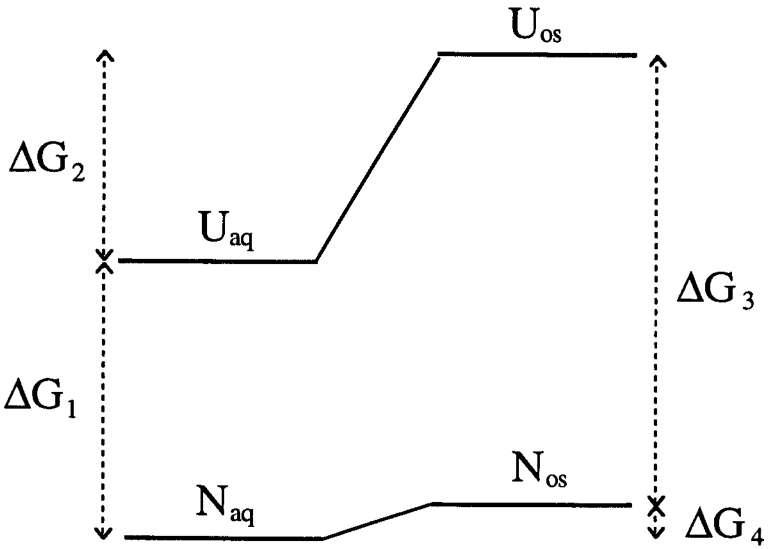


Fig. 1. (Reprinted with permission from **ref.** ***28***.)

measurements ***(13)***. In fact, Timasheff and colleagues ***(30)*** have clearly shown the thermodynamic relationship between transfer Gibbs energy and the preferential interaction parameters measured and used in their approach. **Figure 1** identifies four processes represented by the unfolding reaction of native protein in aqueous solution to unfolded protein in aqueous solution (ΔG°_1); the transfer of unfolded protein in aqueous solution to unfolded protein in (for example) 1 *M* protecting osmolyte solution (ΔG°_2); the unfolding reaction of native protein in 1 *M* osmolyte solution to unfolded protein in 1 *M* protecting osmolyte solution (ΔG°_3); and the transfer of native protein in aqueous solution to native protein in 1 *M* protecting osmolyte solution (ΔG°_4). These events comprise the thermodynamic cycle presented in **Scheme 1**.

Several conclusions can be drawn from **Fig. 1**. First, the stabilization by protecting osmolyte does not arise from osmolyte stabilization of the native state. In fact, the Gibbs energy diagram shows that transferring native protein from water to 1 *M* protecting osmolyte solution increases the Gibbs energy, meaning that the (ΔG°_4) transfer has a positive sign and transfer is therefore desta-

$$\begin{tikzcd}[row sep=huge, column sep=huge]
N_w \arrow[r, shift left, "\Delta G_1"] \arrow[d, shift right, "\Delta G_4"'] & U_w \arrow[l, shift left] \arrow[d, shift left, "\Delta G_2"] \\
N_{sol} \arrow[r, shift right, "\Delta G_3"'] \arrow[u, shift right] & U_{sol} \arrow[l, shift right] \arrow[u, shift left]
\end{tikzcd}$$

Scheme 1

bilizing ***(27)***. Second, the Gibbs energy of transfer for the unfolded state of the protein from aqueous solution to 1 *M* osmolyte ($\Delta G°_2$) is increased much more than ($\Delta G°_4$), suggesting that exposure of the protein fabric to osmolyte solution is a solvophobic process. Third, protein unfolding in 1 *M* osmolyte solution is more unfavorable than unfolding in aqueous solutions ($\Delta G°_3$ is more positive than $\Delta G°_1$). This is a direct consequence of the relatively large unfavorable transfer of unfolded protein from water to 1 *M* osmolyte solution ***(13,26–29)***. Thus, osmolyte stabilization of proteins arises principally from the destabilization of the unfolded state of the protein in the presence of osmolyte. The importance of this mechanism is that osmolytes act primarily on the denatured state in bringing about stabilization, while leaving the native state relatively unaffected and functional (through preferential exclusion of osmolyte). This mechanism provides the means for satisfying the two conditions for natural selection—i.e., stabilizing proteins against the denaturing stress—while not interfering with the functional activity of macromolecules.

4. Transfer Gibbs Energy

From **Fig. 1**, it is clear that the (unfavorable) solvophobic effect of transferring unfolded protein from water to 1 *M* osmolyte solution is critical to understanding the chemical origin of the stabilization of proteins by protecting osmolytes. Thus, identification of the part or parts of the protein responsible for the solvophobic effect is extremely helpful in identifying the nature of the force or forces responsible. The method best suited for identifying the relative favorability or unfavorability of transferring protein functional groups from water to osmolyte solutions involves the measurements of transfer Gibbs energies ($\Delta g°_{tr}$). $\Delta g°_{tr}$ gives a measure of the preference of that functional group for interaction with the cosolvent species relative to its preference for interaction with water ***(31,32)***. Transfer Gibbs energy measurements have a long and storied history in providing an understanding of how strong denaturants such as urea and guanidinium chloride work their magic, and in identifying and characterizing such important fundamental thermodynamic forces as the hydrophobic effect ***(31–33)***. For the purpose of understanding how protecting osmolytes work, transfer Gibbs energy changes are the measurements of choice.

4.1. Theory Behind Transfer Gibbs Energy Measurements

First, a word about terminology. Let's say we want to determine whether some particular functional group on the protein—for example, the peptide backbone—contributes to the solvophobic effect observed on transfer of an unfolded protein from water to 1 *M* sucrose. We then evaluate the Gibbs energy of transfer of the peptide backbone from water to 1 *M* sucrose. The sign of the Gibbs energy change on transfer ($\Delta g°_{tr}$) of the peptide backbone from water to 1 *M* sarcosine tells us with which of the two solvent components the backbone would rather be in contact. And the magnitude of the $\Delta g°_{tr}$ tells us by how much (in energy terms) the backbone prefers the favored solvent. It is important to note that water is always taken to be the reference-state solvent, and an unfavorable Δg°_{tr} means that when given a choice of the two solvents, the peptide backbone prefers to interact with water more than it does with 1 *M* sucrose. Later in this chapter, we will find that the Δg°_{tr} of the peptide backbone from water to protecting osmolyte solutions is indeed unfavorable, and from the positive sign of Δg°_{tr}, it is common to say that the osmolyte interacts unfavorably with the peptide backbone. This statement means that relative to the interaction of water with the peptide backbone, osmolyte solution interacts unfavorably with the backbone. The transfer of peptide backbone from water to protecting osmolyte solution is a solvophobic effect—it is solvophobic with respect to water as the reference solvent ***(28)***.

The application of transfer Gibbs energy measurements to proteins was advanced in development in the 1930s ***(34,35)***. There are many variations of the method, and the one we have used is based on the solubility measurements of amino acids in water and in the specified osmolyte solution. **Fig. 2** shows solubility plots of alanine in water and alanine in 1 *M* TMAO ***(27)***. The limits of solubility of alanine in water and in 1 *M* TMAO are represented by the alanine concentrations at the intersections of the two lines in each solubility plot. At these solubility limits, the chemical potential of the amino acid in the crystal is equal to the chemical potential of the amino acid in the solvent. Thus, for the two experiments we can write:

$$\mu(\text{AA crystal}) = \mu(\text{AA in water}) \tag{1}$$

$$\mu(\text{AA crystal}) = \mu(\text{AA in 1 } M \text{ TMAO}) \tag{2}$$

Because the chemical potential of the amino acid in the crystal is the same in both experiments we can write:

$$\mu(\text{AA in water}) = \mu(\text{AA in 1 } M \text{ TMAO}) \tag{3}$$

And this equation can be expanded in the usual manner to give:

$$\mu^\circ(\text{AA in water}) + RT \ln a(\text{AA in water}) =$$

$$\mu^\circ(\text{AA in 1 } M \text{ TAMO}) + R\,T \ln a(\text{AA in 1 } M \text{ TMAO}) \tag{4}$$

where a(AA in water) and a(AA in 1 *M* TMAO) represent the activities of the amino acid in water and 1 *M* TMAO respectively.

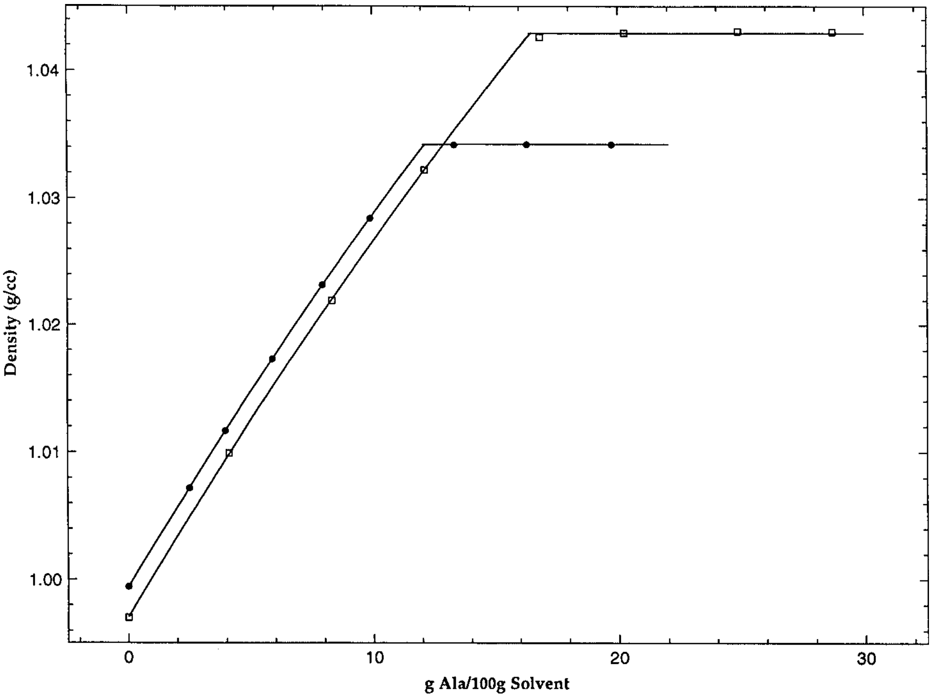


Fig. 2. Density of supernatant solutions vs composition of alanine in water and 1 *M* TMAO solutions. Solubility limits of alanine in water *(open squares)* and in 1 *M* TMAO *(filled circles)* are determined at the intersection points of the solid lines.

Collecting terms in the above equation gives:

$$\begin{gathered}\Delta\mu^\circ = R\,T \ln [a(\text{AA in 1 } M \text{ TMAO})/a(\text{AA in water})] \\ \text{where } \Delta\mu^\circ = \mu^\circ(\text{AA in 1 } M \text{ TMAO}) - \mu^\circ(\text{AA in water})\end{gathered} \qquad (5)$$

It is very difficult to evaluate the activities of the individual components of a three-component mixture, so activities are commonly taken as molar concentrations and the equation becomes:

$$\Delta G^\circ_{tr} = \Delta\mu^\circ = R\,T \ln [C(\text{AA in 1 M TMAO})/C(\text{AA in water})]$$

where C(AA in 1 *M* TAMO) is the molar concentration of the amino acid at the solubility limit in 1 *M* TMAO, C(AA in water) is the amino-acid (molar) concentration at the solubility limit in water, and ΔG°_{tr} is the standard transfer Gibbs energy of the amino acid from water to 1 *M* TMAO ***(29,36,37)***.

4.2. Methods

Transfer Gibbs energy values for 20 common amino acids can be evaluated from solubility measurements such as those shown in **Fig. 2**. From the tabulated values of ΔG°_{tr}, the Gibbs energy of transfer of side chains (Δg°_{tr}) may be estimated by subtracting ΔG°_{tr} of glycine from all other amino acids ***(32,38)***. To determine Δg°_{tr} for the peptide backbone, the solubilities of the cyclic glycylglycine model compound, diketopiperazine (DKP) can be evaluated in water and in a solution of fixed osmolyte concentration. The resulting transfer Gibbs energy of DKP from water to fixed osmolyte concentration is divided by two in order to express the transfer Gibbs energy on a per mol peptide backbone basis ***(27)***. The tabulated Δg°_{tr} values for the side chains and the peptide backbone serve as the database for evaluating the effects of osmolyte on protein denaturation.

5. The Transfer Model and Protein Folding

5.1. What is Being Modeled?

Scheme 1 succinctly outlines the relationship between transfer and protein stability, with ΔG°_4 and ΔG°_2 representing the Gibbs energies of transfer of native state and denatured state from water to cosolvent solution, respectively. These transfer processes are related to the unfolding Gibbs energy changes in water (ΔG°_1) and in osmolyte solution (ΔG°_3) by **Eq. 6**:

$$\Delta G^\circ_2 - \Delta G^\circ_4 \rightleftarrows \Delta G^\circ_3 - \Delta G^\circ_1 \tag{6}$$

which shows that by determining the transfer of native state (ΔG°_4) or denatured state (ΔG°_2) from water to osmolyte solution, one can obtain a measure of the increase in the protein's stability in osmolyte over its stability in water ($\Delta G^\circ_3 - \Delta G^\circ_1$).

By means of preferential interaction measurements, Lin and Timasheff ***(30)*** experimentally determined ΔG°_2 and ΔG°_4 for ribonuclease T1. But another way of obtaining $\Delta G°_2$ and $\Delta G°_4$ is to estimate these quantities by use of models of the denatured and native protein. The latter approach, called the Transfer Model, was used by Nozaki and Tanford to obtain insight into how strong cosolvent denaturants such as urea and guanidine hydrochloride destabilize proteins ***(32,38–40)***. The advantage of using the Transfer Model is that it identifies and quantifies which functional groups on the native and denatured states of the protein being transferred from water to cosolvent solution are principally responsible for the differences in protein stability in the cosolvent solution relative to the protein's stability in water. Much of what we know about the efficacies of urea and guanidine hydrochloride and their interactions with protein functional groups originates from Nozaki and Tanford's pioneering use of the Transfer Model to understand urea and guanidine hydrochloride denaturation ***(38)***. In this chapter, we have adopted the Transfer Model for the purpose of understanding the energetics of interactions between osmolytes and groups on the protein that protect proteins against denaturation.

Operationally, the Transfer Model consists of identifying the solvent-accessible functional groups on the native and denatured forms of the protein; quantifying the fraction of solvent-exposed surface area; multiplying the fractions of solvent-exposed functional groups by their Gibbs energies of transfer; and summing all of the transfer Gibbs energy contributions to obtain estimates of the transfer Gibbs energy of the native ($\Delta G°_4$) or denatured species ($\Delta G°_2$) from water to the specific cosolvent solution. The last step requires appropriate models for the denatured and native protein species, and appropriate models for determining the Gibbs energy of transfer of side chain and backbone functional groups. Such models are described in **Subheading 5.2.**

5.2. Assumptions of the Transfer Model

Amino-acid transfer Gibbs energy measurements have had a long and important history, and they have played a key role in understanding the effect of denaturants and organic solvents on protein stability ***(31–35,38–44)***. The basic idea for use of these quantities is that the sum of the Gibbs energy of transfer of the solvent-exposed component parts of a protein from water to cosolvent equals the Gibbs energy of transfer of the whole protein. This view is based on the assumption that the contributions of component parts of the protein are additive, and this assumption is one of the weaknesses of the model. Several investigators have struggled with the question of additivity with much, but not all, of the data in support of the concept of additivity of side-chain contributions ***(32,38,40,45–50)***. Similarly, additivity of the peptide backbone unit appears to depend on chain length, and while some have presented data indicating the transfer Gibbs energy of peptide backbone unit is additive, others suggest nonadditivity ***(32,38,40,46,47)***. Because of the long history of the issue of additivity, it is doubtful that a convincing case, could settle the issue without extensive studies using a wide variety of compounds and solvent systems. Our view is that additivity for side chain and backbone units occurs to a significant extent—or the ability to obtain reasonable unfolding Gibbs energy changes based upon side-chain and backbone-transfer free-energy data would not be as successful as it continues to be ***(43,44)***. It should be noted that additivity of side-chain transfer free-energy data of Nozaki and Tanford is incorporated as an essential part of theories that have been successful in modeling the thermodynamics of protein folding ***(51)***.

In addition to additivity, several other assumptions and approximations affect the magnitude of Gibbs energies obtained from application of the Transfer Model. As previously mentioned, activity coefficients for three-component systems are so difficult to evaluate that we have little choice but to use concentrations rather than activities to determine transfer Gibbs energies ***(36,37,52)***. Also, Nozaki and Tanford have discussed the issue of what constitutes a good model of the peptide backbone, and several different models have been used to evaluate the transfer Gibbs energy of the peptide backbone ***(33,39,40,42)***. The model we use—diketopiperazine (DKP, alternatively known as cyclic glycylglycine)—is found to give transfer Gibbs energies within 10–15% of those reported for the preferred model used by Nozaki and Tanford ***(27)***. Thus, the transfer Gibbs energy for the peptide backbone may also be a potential source of some error.

Despite the various assumptions and approximations used in the Transfer Model, the Gibbs energy of transfer of native and denatured RNase T1 from water to 1 *M* TMAO obtained from the Transfer Model ($\Delta G°_4$ = 1.7 and $\Delta G°_2$ = 5.9 kcal/mol) agree favorably with the respective values of 1.2 and 5.4 kcal/mol determined experimentally by Lin and Timasheff ***(29,30)***. The agreement provides evidence that the approximations and assumptions are not debilitating to the model, and it also provides a degree of confidence that the major factors responsible for protein stabilization by osmolytes can be identified from transfer data.

Despite the assumptions and approximations of the Transfer Model, it has contributed prominently to the development of the concept of hydrophobic interactions, in understanding denaturant-induced denaturation, and in issues involving protein stabilization ***(27–29,31, 41,49,50,53)***. The strength of the Transfer Model is that from its application, one can identify the major solvent-protein fabric interactions responsible for protein stability. It is important to have a clear understanding of the limitations of the Transfer Model inherent in the assumptions and approximations, and to avoid making too much of it quantitatively. With these conditions, use of the Model provides insight into a number of fundamental principles of protein folding.

5.3. Implementation of the Transfer Model

With tabulated values of side-chain and backbone-unit transfer Gibbs energies from water to osmolyte solutions of fixed concentration, one needs only the crystal coordinates of a protein of interest to transfer the native state from water to cosolvent. Operationally, $\Delta G°_4$ and $\Delta G°_2$ are evaluated according to **Eq. 7**, where $\Delta G°_i$ represents the standard-state transfer Gibbs energy of the peptide backbone units and each of the various amino-acid side chains, n_i is the number of residues of amino acid “i” in the protein, $\Delta G°_i$ is the transfer Gibbs energy of the side chain of residue of type “i”, and α_i is the mean fractional accessibility of the n_i amino-acid side chains or peptide backbone units in the (native or unfolded) protein species of interest ***(27)***.

$$\Delta G°_{2 \text{ or } 4} = \Sigma\, n_i \alpha_i \Delta g°_i \tag{7}$$

To illustrate a calculation of minimal sophistication, let's say we want to evaluate α for the side chain of alanine in the native state of a protein which has a total of 11 alanine residues. Of these, only five have alanine side chains accessible to solvent, each with varying degrees of solvent exposure. The term $\alpha_{\text{(ala side chain)}}$ for the alanine side chains is obtained by summing the solvent accessible surface areas contributed by the five alanine side chains that are partially accessible to solvent. This sum is then divided by a standard-state accessible surface area for the alanine side chain derived from the extended tripeptide gly-ala-gly, with dihedral angles $\Phi = -140°$, $\Psi = 135°$, $X = -120°$ ***(54,55)***.†

The quantity resulting from this interim numerical evaluation ($\alpha^*_{\text{(ala side chain)}}$) represents the number of alanine side chains fully accessible to solvent. Dividing this number by 11 gives the mean (fractional) exposure of alanine side chains accessible to solvent in the native state of the protein, $\alpha_{\text{(ala side chain)}}$. Such calculations are performed for the side chains of all 20 amino acids in the protein plus the peptide backbone, and the α_i values are used in **Eq. 7** to give $\Delta G°_4$, the transfer Gibbs energy of the native state of the protein from water to the osmolyte.

Although a variety of programs exist for evaluating the accessible surface area (α_i values), we have used the static accessible surface area algorithm of Lee and Richards as modified by Lesser & Rose, taking 1.4 Ås as the probe size of the solvent ***(55,56)***. While the crystallographic coordinates of the protein provides a model of the native state of the protein, the lack of defined structure for the denatured state makes it necessary to model the denatured ensemble. Creamer et al. ***(57)*** examined the complexities of the denatured ensemble, and have suggested two limiting models that bracket the expected behavior of an unfolded chain. One limiting model of the denatured state, which sets the upper boundary of solvent accessibility, involves a hard sphere model (with attending excluded volume effects) that behaves as a homopolymer in a good solvent. That is, the model representing the upper boundary is highly solvent-accessible. The limiting model that sets the lower boundary of solvent accessibility is generated from excised fragments of folded proteins that retain intramolecular interactions. This model is representative of a compact denatured ensemble. **Equations 8** and **9** provide the means to estimate the Gibbs energy changes for transfer of

$$\Delta G^\circ_{tr\ lb} = \Sigma\Delta g^\circ_i * n_i * \alpha_i * (\text{lower ASA/stand ASA}) \tag{8}$$

$$\Delta G^\circ_{tr\ ub} = \Sigma\Delta g^\circ_i * n_i * \alpha_i * (\text{upper ASA/stand ASA}) \tag{9}$$

the highly solvent accessible (upper boundary) denatured state model ($\Delta G^\circ_{tr\ ub}$) and the compact (lower boundary) denatured state model ($\Delta G^\circ_{tr\ lb}$) from water to osmolyte solution ***(28)***. In these calculations, Δg°_i and n_i have the same meanings as in **Eq. 7**. However, α_i values are obtained by first taking all side-chain and backbone units as being fully solvent accessible (i.e., α^* as defined above is set to unity), and the mean fractional solvent exposure of side-chain or backbone unit "i" is obtained from $\alpha_i = \alpha^*/n_i$. Finally, (stand ASA) represents the standard accessible surface area of each type of amino-acid side chain or backbone U calculated using a stochastic standard-state model for gly-X-gly[†], while (lower ASA) and (upper ASA) are the lower and upper boundary accessible surface areas of each particular side chain and backbone as given in **Table 1** of Creamer et al. ***(57)***. Thus, transfer Gibbs energies of denatured ensembles that are highly solvent-accessible, those that have low solvent accessibilities (compact denatured ensemble), and those in between are bracketed by $\Delta G^\circ_{tr\ ub}$ and $\Delta G^\circ_{tr\ lb}$.

6. Conclusion

When used with knowledge of its shortcomings, the Transfer Model has been highly successful in providing fundamental under-

---

[†] Two models are used in these calculations: the extended tripeptide model, gly-X-gly, and the stochastic gly-X-gly model ***(54,55)***.

standing and insight into such basic processes as the hydrophobic effect, urea and guanidine hydrochloride denaturation of proteins, and the stabilization of proteins by naturally occurring osmolytes. This model is particularly helpful in identifying the major component parts of the protein that—through interaction with solvent—result in the folded structure that is the native state. The procedures and bases for transfer Gibbs energy changes are elementary in their conception, and are also steeped in the history of biophysical chemistry.

References

1. Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., and Somero, G. N. (1982) Living with water stress: evolution of osmolyte systems. *Science* **217,** 1214–1222.
2. Stewart, J. A. and Ouellet, L. (1959) *Can. J. Chem.* **37,** 744–750.
3. Brown, A. D. and Simpson, J. R. (1972) Water relations of sugar-tolerant yeasts: the role of intracellular polyols. *J. Gen. Microbiol.* **72,** 589–591.
4. Pollard, A. and Wyn Jones, R. G. (1979) *Planta.* **144,** 291–298.
5. Borowitzka, L. J. and Brown, A. D. (1974) The salt relations of marine and halophilic species of the unicellular green alga, Dunaliella. The role of glycerol as a compatible solute. *Arch. Microbiol.* **96,** 37–52.
6. Roser, B. (1991) *BioPharm.* **4,** 47–53.
7. Crowe, J. H. (1971) *Am. Nat.* **105,** 563–573.
8. Yancey, P. H. and Somero, G. N. (1980) *J. Exp. Zool.* **212,** 205–213.
9. Somero, G. N. (1986) Protons, osmolytes, and fitness of internal milieu for protein function. *Am. J. Physiol.* **251,** R197–R213.
10. Clark, M. E. (1985) The osmotic role of amino acids discovery amd function, in *Transport Processes, Iono- and Osmoregulation* (Gilles, R. and Gilles-Baillien, M., eds.), Springer-Verlag, Berlin Heidelberg, pp. 412–423.
11. Yancey, P. H. and Somero, G. N. (1979) Counteraction of urea destabilization of protein structure by methylamine osmoregulatory compounds of elasmobranch fishes. *Biochem. J.* **183,** 317–323.
12. Forster, R. P. and Goldstein, L. (1976) Intracellular osmoregulatory role of amino acids and urea in marine elasmobranchs. *Amer. J. Physiol.* **230,** 925–931.
13. Arakawa, T., Bhat, R., and Timasheff, S. (1990) Why preferential hydration does not always stabilize the native structure of globular proteins. *Biochemistry* **29,** 1924–1931.

14. Arakawa, T. and Timasheff, S. N. (1982) Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry* **21,** 6545–6552.
15. Arakawa, T. and Timasheff, S. N. (1982) Stabilization of protein structure by sugars. *Biochemistry* **21,** 6536–6544.
16. Arakawa, T. and Timasheff, S. N. (1983) Preferential interactions of proteins with solvent components in aqueous amino acid solutions. *Arch. Biochem. Biophys.* **224,** 169–177.
17. Arakawa, T. and Timasheff, S. N. (1984) Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry* **23,** 5912–5923.
18. Arakawa, T. and Timasheff, S. N. (1984) Protein stabilization and destabilization by guanidinium salts. *Biochemistry* **23,** 5924–5929.
19. Arakawa, T. and Timasheff, S. N. (1984) The mechanism of action of Na glutamate, lysine HCl, and piperazine-N,N'-bis(2-ethanesulfonic acid) in the stabilization of tubulin and microtubule formation. *J. Biol. Chem.* **259,** 4979–4986.
20. Arakawa, T. and Timasheff, S. N. (1985) The stabilization of proteins by osmolytes. *Biophys. J.* **47,** 411–414.
21. Lee, J. C. and Lee, L. L.-Y. (1981) Preferential solvent interactions between proteins and polyethylene glycols. *J. Biol. Chem.* **256,** 625–631.
22. Lee, L. L.-Y. and Lee, J. C. (1987) Thermal stability of proteins in the presence of poly(ethylene glycols). *Biochemistry* **26,** 7813–7819.
23. Timasheff, S. N. (1993) The control of protein stability and association by weak interactions with water: how do solvents affect these processes? *Annu. Rev. Biophys. Biomol. Struct.* **22,** 67–97.
24. Timasheff, S. N. (1994) Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry* **33,** 12,695–12,701.
25. Lee, J. C. and Timasheff, S. N. (1981) The stabilization of proteins by sucrose. *J. Biol. Chem.* **256,** 7193–7201.
26. Arakawa, T., Bhat, R., and Timasheff, S. (1990) Preferential interactions determine protein solubility in three-component solutions: the MgCl2 system. *Biochemistry* **29,** 1914–1923.
27. Liu, Y. and Bolen, D. W. (1995) The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry* **34,** 12,884–12,891.

28. Qu, Y., Bolen, C. L., and Bolen, D. W. (1998) Osmolyte-driven contraction of a random coil protein. *Proc. Natl. Acad. Sci. USA* **95,** 9268–9273.
29. Wang, A. and Bolen, D. W. (1997) A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry* **36,** 9101–9108
30. Lin, T.-Y. and Timasheff, S. N. (1994) Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein. *Biochemistry* **33,** 12,695–12,701.
31. Cohn, E. J. and Edsall, J. T. (1943) Interaction between organic solvents and dipolar ions estimated from solubility ratios, in *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions*, Reinhold Publishing Corp., New York, NY.
32. Nozaki, Y. and Tanford, C. (1963) *J. Biol. Chem.* **238,** 4074–4080.
33. Robinson, D. R. and Jencks, W. P. (1965) *J. Am. Chem. Soc.* **87,** 2462–2470
34. McMeekin, T. L., Cohn, E. J., and Weare, J. H. (1935) *J. Am. Chem. Soc.* **57,** 626–633.
35. McMeekin, T. L., Cohn, E. J., and Weare, J. H. (1936) *J. Amer. Chem. Soc.* **58,** 2173–2181.
36. Uedaira, H. (1972) *Bul. Chem. Soc. Jpn.* **45,** 3068–3072.
37. Uedaira, H. (1977) *Bul. Chem. Soc. Jpn.* **50,** 1298–1304.
38. Nozaki, Y. and Tanford, C. (1965) The solubility of amino acids and related compounds in aqueous thylene glycol solutions. *J. Biol. Chem.* **240,** 3568–3573.
39. Nozaki, Y. and Tanford, C. (1970) The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **245,** 1648–1652.
40. Nozaki, Y. and Tanford, C. (1971) The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **246,** 2211–2217.
41. Tanford, C. (1970) Protein denaturation. Theoretical models for the mechanism of denaturation, in *Adv. Protein Chem.* vol. 24, Academic Press, New York, NY, pp. 1–95.
42. Robinson, D. R. and Jencks, W. P. (1965) *J. Am. Chem. Soc.* **87,** 2470–2479.
43. Staniforth, R. A., Burston, S. G., Smith, C. J., Jackson, G. S., Badcoe, I. G., Atkinson, T., et al. (1993) The energetics and cooperativity of

protein folding: a simple experimental analysis based upon the solvation of internal residues. *Biochemistry* **32,** 3842–3851.

44. Pace, C. N. (1975) The stability of globular proteins. *CRC Crit. Rev. Biochem.* **3,** 1–43.
45. Edsall, J. T. and Wyman, J. (1958) *Biophysical Chemistry,* I, Academic Press Inc., London.
46. Lapanje, S., Skerjanc, J., Glavnik, S., and Zibret, S. (1978) *J. Chem. Thermodynamics* **10,** 425–433.
47. Nandi, P. K. and Robinson, D. R. (1984) Effects of urea and guanidine hydrochloride on peptide and nonpolar groups. *Biochemistry* **23,** 6661–6668.
48. Schrier, M. Y. and Schrier, E. E. (1976) Transfer free energies and average static accessibilities for ribonuclease A in guanidinium hydrochloride and urea solutions. *Biochemistry* **15,** 2607–2612.
49. Tanford, C. (1964) *J. Am. Chem. Soc.* **86,** 2050–2059.
50. Wetlaufer, D. B., Malik, S. K., Stoller, L., and Coffin, R. L. (1964) *J. Am. Chem. Soc.* **86,** 508–514.
51. Alonso, D. and Dill, K. (1991) Solvent denaturation and stabilization of globular proteins. *Biochemistry* **30,** 5974–5985.
52. Schönert, H. and Stroth, L. (1981) *Biopolymers* **20,** 817–831.
53. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, in *Adv. in Prot. Chem.* vol. 14, Academic Press, New York, NY, pp. 1–63.
54. Chothia, C. (1975) *J. Mol. Biol.* **105,** 1–14.
55. Lesser, G. J. and Rose, G. D. (1990) Hydrophobicity of amino acid subgroups in proteins. *Proteins Struct. Funct. Genet.* **8,** 6–13.
56. Lee, B. and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400.
57. Creamer, T. P., Srinivasan, R., and Rose, G. D. (1997) Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* **36,** 2832–2835.

3

The Thermodynamic Linkage Between Protein Structure, Stability, and Function

Ernesto Freire

1. Introduction

For many years, the ability to predict protein stability or protein function from structure has been considered a major scientific goal with significant practical implications in protein engineering. The Human Genome Project has accelerated the need to achieve that goal. With nearly 100,000 protein-coding genes, fast computer algorithms will be required for structural determination as well as for physical and functional analysis. With more than 10,000 structures already in the protein database, structure alone is clearly not sufficient for a comprehensive understanding of function and regulation. Human intervention in medicine and biotechnology—either by genetic manipulation aimed at introducing specific protein mutations that modify stability or functionality, or by designing new drugs that regulate or modify protein function—requires a clear understanding of the relationships between structure and function. These relationships are mediated by energetics. Proteins are closely packed macromolecules exhibiting a large number of noncovalent atomic interactions that define not only their three-dimensional structure and stability, but also their functional properties. While

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

the relationships between energetics, folding, and stability have been explored by many researchers, the relationships between protein stability and functional cooperativity remain largely unknown. It is well-known that many proteins undergo conformational changes during function or that conformational flexibility is critical to biological activity. These functional properties are dictated by the same interactions that define stability. Thus, functional properties of proteins (e.g., the transmission of binding effects, cooperative interactions, molecular signaling, and allosterism) are defined by energetics, and more precisely by the specific way in which stabilizing interactions are distributed throughout the protein structure. These interactions not only define the overall structure and stability of a protein, but the presence and location of regions with different propensities to undergo conformational rearrangements. Stability and function are linked to each other, and the elucidation of the specific ways in which they are related is a major goal in the postgenomic era.

The development of algorithms for structure-based stability and function prediction requires accurate and rapid parametric equations that relate structure to energetics. Ideally, these equations must be derived from protein and peptide data in aqueous solution rather than small organic molecules. A great deal of experimental work is being done on the solution thermodynamics of peptides. The rapid growth in the number of high-resolution protein structures, combined with the increasing number of proteins for which high-resolution thermodynamic data is also available, has provided a database for the development of empirical functions that relate thermodynamic and structural parameters. Using this approach, several research groups have developed structural parameterizations of the various thermodynamic parameters that define the Gibbs energy (the enthalpy, entropy, and heat-capacity changes) ***(1–21)***. This approach has been shown to be accurate and has permitted the exploration of different stability and functional properties of proteins. This chapter summarizes recent developments from this laboratory in the area of structure-based prediction of protein stability and functional cooperativity.

2. The Stability of Proteins

Discussions about protein stability are usually limited to the stability of the native state against denaturation. Yet the native state may include different functionally relevant conformations characterized by different Gibbs energies and therefore different stabilities (e.g., the R and T states of hemoglobin). In many cases the conformation of a protein changes when a ligand is bound (e.g., the HIV protease); in this case, the two conformations have different Gibbs energies and therefore different intrinsic stabilities. Even when the native state does not undergo a conformational change, it is still characterized by the occurrence of a large number of local unfolding events that give rise to many substates. Thus, the native state itself needs to be considered as a statistical ensemble of conformations rather than a unique entity. These distinctions are very important from a functional point of view since different conformations are usually characterized by different functional properties. It should be clear from this discussion that protein redesign for stability purposes must consider explicitly the effects of the introduced mutations on the energetic differences between states and their functional repercussions.

3. The Stability of the Native State

The stability of the native state (and of any other conformational state) of a protein is dictated by the magnitude of its Gibbs energy (ΔG), which is given by the standard thermodynamic expression:

$$\Delta G = \Delta H - T\Delta S \tag{1}$$

where ΔH and ΔS are the enthalpy and entropy changes respectively. In this chapter, we will follow the common practice of using the native state as the reference state. It was observed many years ago (*see* **refs. *22–24***) that both the enthalpy and entropy changes were not constants but increasing functions of temperature, and that the Gibbs energy of stabilization of a protein needs to be written as:

$$\Delta G = \Delta H(T_R) + \Delta C_p \cdot (T - T_R) - T[\Delta S(T_R) + \Delta C_p \cdot \ln(T/T_R)] \tag{2}$$

where T_R is a convenient reference temperature, ΔC_p is the heat capacity change, and $\Delta H(T_R)$ and ($\Delta S(T_R)$ are the enthalpy and entropy values at that temperature. Because 60°C (333.15 K) is close to the median denaturation temperature of proteins, it has been chosen as a convenient reference temperature, and will be used in this chapter unless noted otherwise. The temperature dependence of ΔH and ΔS (i.e., ΔC_p) is also an important parameter because it transforms the Gibbs energy function from a linear into a parabolic function of temperature, as illustrated in **Fig. 1A**. The midpoint of the temperature denaturation of a monomeric protein (T_m) occurs when $\Delta G = 0$. It is evident from the curves in **Fig. 1** that for large values of ΔC_p, the Gibbs energy crosses zero twice—one at high temperature (heat denaturation) and one at low temperature (cold denaturation). The native state is thermodynamically stable between those two temperatures and ΔG exhibits a maximum at the temperature at which $\Delta S = 0$. For extended temperature extrapolations, the temperature dependence of ΔC_p must be considered ***(15)***. For the stability of oligomeric proteins, the reader is referred to **refs.** ***25–27***.

The peculiar shape of the Gibbs energy function of a protein does not permit a unique definition of protein stability. For example, having a higher denaturation temperature does not necessarily imply that a protein will be more stable at room temperature or at a physiological temperature, as illustrated in **Fig. 1B**. It is perfectly feasible for a protein to have a higher denaturation temperature than another protein, but a lower stability at room temperature.

The temperature stability of a protein is fully specified if three thermodynamic parameters are known—the enthalpy and entropy changes at the reference temperature and the heat capacity change (ΔH, ΔS, and ΔC_p). These parameters can be measured by high-sensitivity calorimetric techniques ***(28,29)***. These are also the parameters that need to be evaluated from structure for accurate stability prediction.

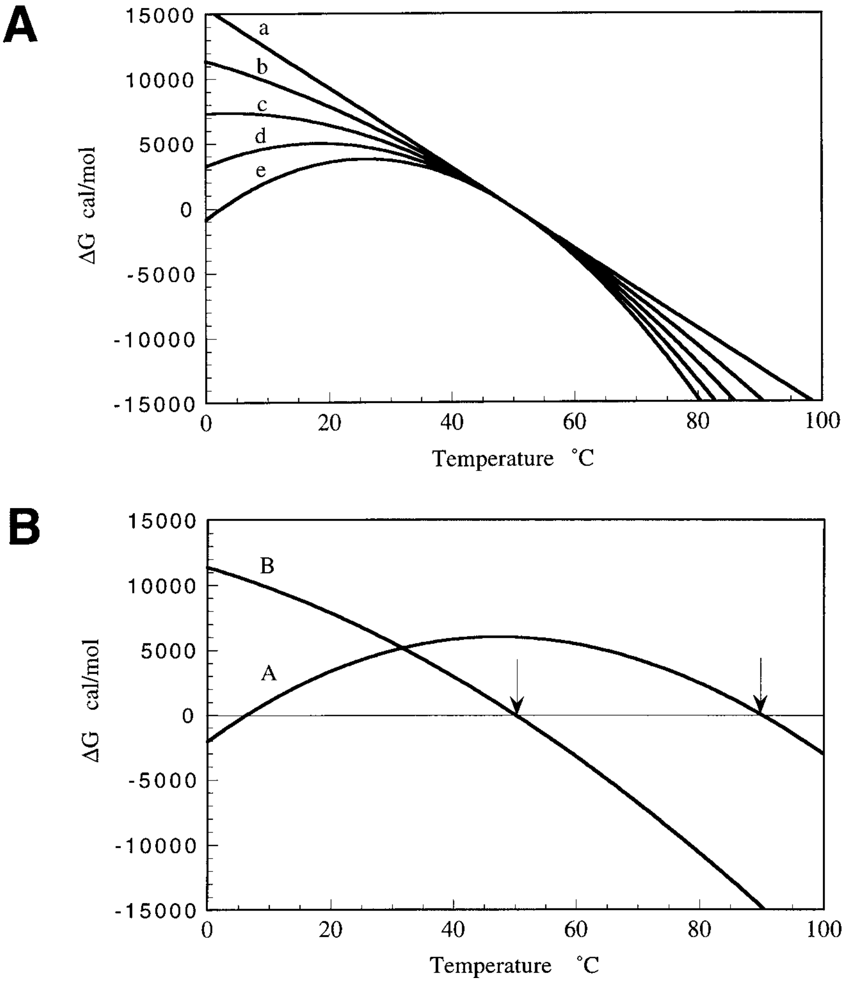


Fig. 1. **(A)** The temperature dependence of the Gibbs energy for different values of the heat-capacity change. In all cases, ΔG is zero at 50°C with an enthalpy change of 100 kcal/mol. The curves correspond to ΔC_p values of: a) zero, b) 1000, c) 2000, d) 3000, and e) 4000 cal/K·mol. **(B)** The temperature dependence of two hypothetical proteins is presented. Although protein A has a higher temperature stability than protein B, protein B is more stable at low temperatures. This example demonstrates that protein stability measured at room temperature (e.g., by urea or GuHCl denaturation) cannot be used as a predictor of high temperature stability.

4. A Realistic Gibbs Energy Function for Computational Analysis

A realistic energy function with the ability to account for the temperature dependence of the Gibbs energy requires the structural prediction of ΔH, ΔS, and ΔC_p. Since these parameters reflect the balance of forces that determine the stability of a protein in a different manner, a comparison between experimental and calculated ΔH, ΔS, and ΔC_p values can be used to evaluate whether the structural parameterization captures the experimental energetic balance correctly. Let us consider hen egg-white lysozyme (HEWL) as an example. This protein is characterized by a Gibbs energy of stabilization close to 7 kcal/mol at 25°C and pH 2.0 ***(21)***. The structure of this protein is known at high resolution, and therefore structure-based stability calculations are feasible. If a computer program could correctly predict ΔG at 25°C, it would be impossible to say anything about the temperature denaturation of that protein or its stability at physiological temperature unless ΔH, ΔS, and ΔC_p were also predicted correctly. Moreover, it may also be possible for a computer program to correctly predict the Gibbs energy of stabilization at some temperature, but for the wrong reasons. Since $\Delta G = \Delta H - T\Delta S$ (*see* **Eqs. 1** and **2**) many combinations of ΔH, ΔS, and ΔC_p values can give the same ΔG. However, not all combinations reflect the actual balance of forces existing in the protein molecule. A comparison of experimental and calculated ΔH, ΔS, and ΔC_p values provides a simple way to validate structure-based calculations. In most laboratories, the redesign of a protein for stability is usually done with the wild-type or another known variant of known structure as a starting point. Thus, we recommend that any computational approach be tested first by comparing ΔG, ΔH, ΔS, and ΔC_p values calculated from the structure of the wild-type with those experimentally measured by high-sensitivity calorimetric techniques. If the computational procedure cannot correctly predict those thermodynamic parameters, the procedure and its energy function will be useless for the design of temperature-stable mutants.

5. The Structural Parameterization of the Energetics

Proteins fold in an aqueous environment. The formation of intramolecular interactions between different chemical groups in a protein is coupled to the disruption of interactions of those same groups with the solvent (water). For those reasons, the energetic parameters associated with protein stability are the differences between the intramolecular interactions that are established when a protein folds and those existing with the solvent (water) when the protein is unfolded. Traditionally, the energetics of protein stability has been estimated by calculating separately and then adding together the contributions of noncovalent intramolecular interactions calculated by some force field for vacuum conditions, and the hydration contributions (*see* **refs. *30*,*31***). While this approach may be theoretically correct, it is extremely inaccurate. Hydration and intramolecular interactions are very large in magnitude, and cancel each other out to a large extent. For example, hydration and intramolecular contributions to the enthalpy change, taken individually, are on the order of thousands of kcal/mol, but of opposite sign. Experimental enthalpies, on the other hand, are on the order of tens of kcal/mol or less at 25°C. Therefore, small errors in hydration or intramolecular contributions result in errors that are unacceptable for reasonable stability predictions.

It seems apparent that structure-based stability predictions would be significantly improved by eliminating the subtraction of intramolecular and hydration parameters. This can be done by developing a set of effective parameters for intramolecular interactions occurring in aqueous solutions rather than in a vacuum. This approach has been followed in our laboratory for the last eight years, and has resulted in a comprehensive structural parameterization of the folding and binding energetics of proteins. The following references contain detailed descriptions of the structural parameterization of the energetics: ***2*,*4–6*,*8–11*,*13–20***. In this chapter, we provide a general discussion and a guide to practical applications of this approach in protein engineering.

An important component of the structural parameterization of the energetics relies on the existence of a joint database of high-resolution structures and high-resolution thermodynamic data. In 1990, when this project was started, there were only 1,000 structures in the protein database and about 10 systems for which high-resolution thermodynamics were available. Today, there are 10,000 structures and about 100 proteins and complexes for which high-quality thermodynamic data are available. In addition, many systems have been measured under different conditions (such as pH, ligands, and salt concentration). This growth in the structure/thermodynamic database has permitted the explicit inclusion of different interactions and a continuous refinement of the basic set of energy coefficients.

Within the context of the structural parameterization of the energetics, the Gibbs energy of stabilization is approximated by the equation:

$$\Delta G = \Delta G_{\mathrm{gen}} + \Delta G_{\mathrm{ion}} + \Delta G_{\mathrm{tr}} + \Delta G_{\mathrm{other}} \tag{3}$$

where ΔG_{gen} contains the contributions typically associated with the formation of secondary and tertiary structure (van der Waals interactions, hydrogen bonding, hydration, and conformational entropy), ΔG_{ion} the electrostatic and ionization effects, and ΔG_{tr} the contribution of the change in translational degrees of freedom existing in oligomeric proteins. The term $\Delta G_{\mathrm{other}}$ includes interactions unique to specific proteins that cannot be classified in a general way (e.g., prosthetic groups, metals, and ligands) and must be treated on a case-by-case basis.

6. The Enthalpy Change

Upon protein unfolding, the major contribution to the enthalpy change arises from the disruption of intramolecular interactions such as van der Waals and hydrogen bonds, and the parallel solvation of the interacting groups. This term is called the generic enthalpy change, ΔH_{gen}, because it involves interactions that are common to all proteins. The atomic contributions to this term are a function of the interactions of each atom within the protein and to the changes in the degree of solvation upon unfolding. At the reference temperature of 60°C, ΔH_{gen} can be written as:

$$\Delta H_{gen}(60) = \sum_i \alpha_i(\bar{\rho}) \cdot \Delta ASA_i \qquad (4)$$

where the summation runs over all atoms in the protein, ΔASA_i is the change in solvent-accessible surface area of atom i upon unfolding, and $\alpha_i(\bar{\rho})$ is an elementary coefficient that depends on the atom type and the average atomic packing density of that atom within the protein. $\alpha_i(\bar{\rho})$ is expressed as a linear function of the number of atoms within van der Waals distance. Other effects also contribute to the enthalpy change, such as the release of prosthetic groups or ligands upon unfolding or changes in the degree of protonation of groups that exhibit anomalous pK_a in the native state of the protein.

The accuracy of the structural parameterization is demonstrated in **Fig. 2**. In this figure, a correlation plot between the experimental and calculated enthalpy values for 44 proteins is shown. The slope of this graph is 1.015, the correlation coefficient is 0.944, and the standard error of the calculated values is ±2.6 kcal/mol. On a percentage basis, the mean error of the structure-based enthalpy prediction is ±3.4%.

7. The Heat Capacity Change

The heat capacity change observed in protein unfolding is largely the result of changes in hydration of groups that are buried from the solvent in the native state, which become solvent exposed upon unfolding. It has been known for many years that changes in heat capacity associated with hydration are local in nature and scale, in terms of the changes in the surface area accessible to the solvent ***(32–34)***. The heat-capacity change is a weak function of temperature and has been parameterized in terms of changes in solvent-accessible surface areas (ΔASA) since it originates mainly from changes in hydration ***(6,15,34)***. The contribution to the heat capacity caused by changes in hydration can be written as:

$$\Delta C_p = \Sigma\, a_i(T) \cdot \Delta ASA_i \qquad (5)$$

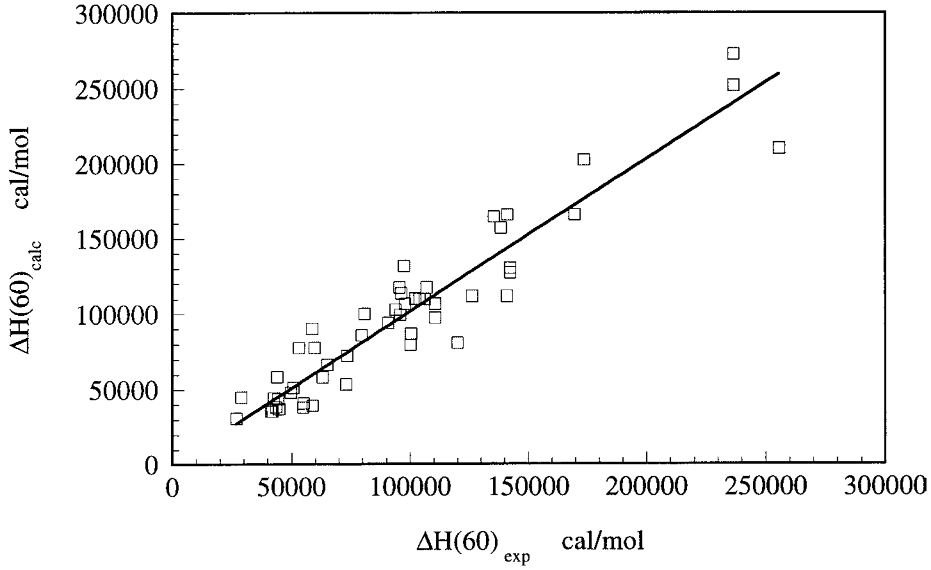


Fig. 2. Correlation plot for the experimental and predicted enthalpy changes for 44 different proteins at the reference temperature of 60°C. The slope of this graph *(solid line)* is 1.015, the correlation coefficient is 0.944, and the standard error of the calculated values is ± 2.6 kcal/mol. On a percentage basis, the mean error of the structure-based enthalpy prediction is ± 3.4%. The experimental values were taken from the compilation made by Robertson and Murphy ***(21)***, and correspond to α-chymotrypsin, α-chymotrypsinogen, α-lactalbumin, acyl carrier protein, arc repressor, B1 of protein G, B2 of protein G, barnase, barstar, BPTI, carbonic anhydrase B, CI2, cyt b5 (tryp frag), cytochrome c (horse), cytochrome c (yeast), GCN4, HIV protease, HPr, IL-1, lac repressor, lysozyme (human), lysozyme (apo equine), lysozyme (holo equine), lysozyme (hen), lysozyme T4, met repressor, myoglobin (horse), myoglobin (whale), OMTKY3, papain, parvalbumin, pepsin, pepsinogen, plasminogen K4 domain, RNase T1, RNase A, Sac7d, SH3 spectrin, Staphylococcus nuclease, stefin A, tendamistat, thioredoxin, trp repressor, and ubiquitin.

where the summation runs over all atoms in the protein. $a_i(T)$ is the contribution of atom *i* per unit area and ΔASA_i is the change in accessible surface area for atom *i*. From studies in model compounds, it has been possible to identify three different elementary contributions. Nonpolar atoms (i.e., carbons) make a positive contribution to ΔC_p when they become hydrated, and polar atoms make a negative contribution—except for hydroxyl oxygens, which make a positive contribution. In general, for low-temperature calculations ($T < 80°C$), the set of temperature-independent coefficients is sufficient ***(15,34)***. Unlike the enthalpy coefficients, the coefficients $a_i(T)$ are essentially independent of the packing density of the groups that become solvated upon unfolding. Specific effects such as heat-capacity changes associated with changes in protonation, differential binding of ligands, or denaturants must be considered individually, as described elsewhere ***(15,34)***—but on average they contribute very little to the total folding/unfolding ΔC_p.

The accuracy of the structure-based prediction of heat capacity changes is demonstrated in **Fig. 3**. In this figure, a correlation plot between the experimental and calculated heat capacity changes for 44 proteins is shown. The slope of this graph is 1.16, the correlation coefficient is 0.966, and the standard error of the calculated values is ±63 cal/K·mol. On a percentage basis, the mean error of the structure-based heat capacity prediction is ±3.6%.

8. The Entropy Change

In the calculation of the unfolding entropy change, two primary contributions are included—one resulting from changes in solvation and the other caused by changes in conformational degrees of freedom ($\Delta S_{gen}(T) = \Delta S_{solv}(T) + \Delta S_{conf}$). The entropy of solvation is temperature-dependent, while the conformational entropy is essentially a constant at different temperatures. The entropy of solvation can be written in terms of the heat capacity change if the temperatures at which the apolar and polar hydration entropies are zero ($T^*_{S,ap}$ and $T^*_{S,pol}$) are used as reference temperatures:

$$\Delta S_{solv}(T) = \Delta S_{solv,ap}(T) + \Delta S_{solv,pol}(T) \qquad (6)$$

$$\Delta S_{solv}(T) = \Delta C_{p,ap} \cdot \ln(T/T^*_{S,ap}) + \Delta C_{p,pol} \cdot \ln(T/T^*_{S,pol}) \qquad (7)$$

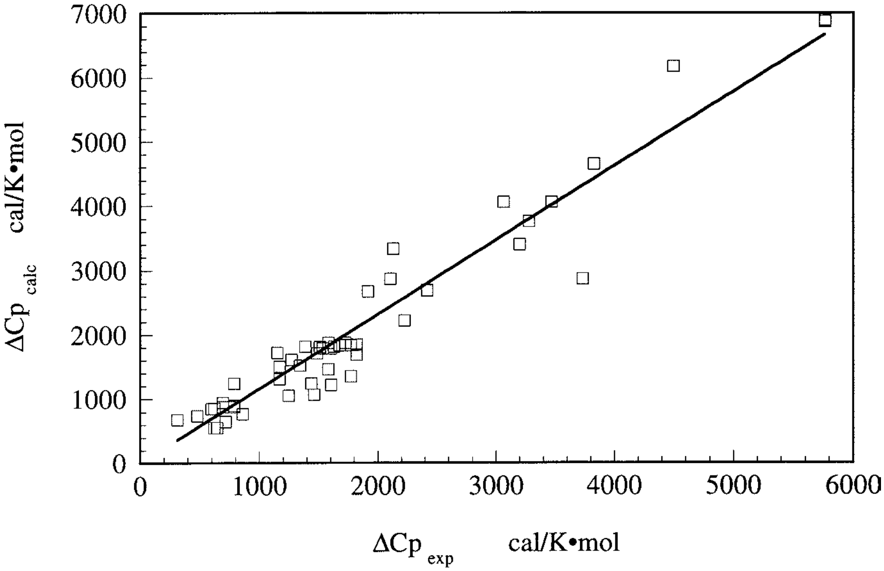


Fig. 3. Correlation plot for the experimental and predicted heat-capacity changes for 44 different proteins. The slope of this graph is 1.16, the correlation coefficient is 0.966, and the standard error of the calculated values is ± 63 cal/K·mol. On a percentage basis, the mean error of the structure-based heat capacity prediction is ± 3.6%. The proteins are the same as in **Fig. 2**.

$T^*_{S,ap}$ has been known to be equal to 385.15 K for some time ***(4,35)***, and $T^*_{S,pol}$ has been recently found to be close to 335.15 K ***(16)***. While the entropy of apolar solvation appears to be additive, the situation for polar solvation is known to depend on the number and proximity of polar functional groups in the molecule ***(36)***. This is reflected in the observation that $T^*_{S,ap}$ is similar for small molecules and proteins, whereas $T^*_{S,pol}$ is an effective parameter for peptides.

In the structural parameterization, conformational entropies are evaluated by explicitly considering the following three contributions for each amino acid: 1) $\Delta S_{bu\text{->}ex}$, the entropy change associated with the transfer of a side chain that is buried in the interior of the protein to its surface; 2) $\Delta S_{ex\text{->}u}$, the entropy change gained by a surface-exposed side chain when the peptide backbone changes from a unique conformation to an unfolded conformation; and 3) ΔS_{bb}, the entropy change gained by the backbone itself upon unfolding from a unique conformation. The magnitude of these terms for each amino acid has been estimated by a computational analysis of the probability of different conformers as a function of the dihedral and torsional angles ***(14,16,18)***. These computations are cpu-intensive, and for this reason the results have been stored in look-up tables so they do not have to be repeated for each calculation. This strategy permits stability computations using only a few seconds of cpu time.

9. Electrostatic and Protonation Contributions

Two types of electrostatic effects as well as protonation effects are included in ΔG_{ion}—Coulombic contributions caused by the interactions between charged sites, and the self-energy term arising from transferring a charge between environments with different dielectric constants. These electrostatic contributions are computed as described by García-Moreno ***(37,38)*** using the standard equation:

$$\Delta G_{ion} = 332\left[\sum_i \frac{Z_i^2}{2r_i}\left[\frac{1}{D} - \frac{1}{D_{ref}}\right] + Z_i \sum_j \frac{Z_j}{D r_{i,j}}\right] \tag{8}$$

where Z is the charge, r_i the radius of the charged particle, $r_{i,j}$ the separation between two charges, D and D_{ref} the dielectric constants in the protein and solvent. Protonation effects are treated as described in **ref.** ***34*** from a knowledge of the pK_a of the groups that change ionization state upon folding/unfolding.

In protein folding, changes in the degree of protonation between conformational states are common. Usually, the folding/unfolding equilibrium is coupled to the protonation/deprotonation of one or more groups in the protein. If this is the case, the protonation contribution to the Gibbs energy of an arbitrary conformational state *j* is equal to:

$$\Delta G_{proton} = -R \cdot T \cdot \sum_j \ln\left[\frac{(1 + 10^{(pK_{a,j} - pH)})}{(1 + 10^{(pK^0_{a,j} - pH)})}\right] \tag{9}$$

where $pK_{a,j}$ is the pK_a of group j in the state under consideration and $pK^0_{a,j}$, the pK of that group in the native state. According to **Eq. 9**, at any pH the additional contribution of protonation to ΔG can be calculated if the pK_a of that group is known in the protein conformations under consideration. Traditionally, the pK_a of different groups in proteins have been estimated experimentally, since computational procedures for the evaluation of pK_as are still inaccurate. It is expected that within the next few years, protonation effects will be accurately parameterized in terms of structure, and that an accurate structure-based prediction of the pH dependence of protein stability will be available.

10. Calculation of Solvent Accessibilities

Within the framework of the structural parameterization, changes in accessible surface areas are calculated by implementing the Lee and Richards algorithm ***(39)***. In all calculations, a solvent radius of 1.4 Å and a slice width of 0.25 Å are used. Because of the nature of the algorithm, different orientations of the macromolecule with respect to the slicing plane yield slightly different solvent accessibilities. These small differences are irrelevant for large ASA changes, such as those associated with the folding/unfolding of proteins, but may be significant in those cases where small differences in accessibilities need to be evaluated (e.g., single amino-acid mutations). In order to better define small changes in solvent accessibility, different molecular orientations (usually 64) with respect to the slicing plane are considered in the accessible surface-area calculations. This procedure provides more robust ASA values because of the many ways in which the molecules are sliced. The resulting ASA for each atom is the numerical average of the values obtained for all molecular orientations.

In the past, solvent accessibilities for unfolded peptides were estimated by using computer-built extended peptides of appropriate sequences, or extended tripeptides of the form Ala-X-Ala or Gly-X-Gly. However, those models are poor descriptors of the unfolded state of a protein, and its use leads to an overestimation of changes in solvent accessibility ***(40)***. In addition, one common characteristic of these computational methods is that they utilize a single static structure. In contrast, the solvent accessibility of a peptide chain in solution is expected to vary with the rotation of side chains, and also in unfolded conformations with backbone rotations. Under equilibrium conditions, the solvent-accessible surface area of an amino acid in an unfolded polypeptide chain is not a constant, but a Boltzmann-weighted average to which all accessible conformations contribute according to their probability. This approach has been utilized in the investigation of the α-helix propensities of all amino acids ***(18)***. From this study, a set of refined values for the solvent accessibilities of different amino acids in the unfolded state was derived. It must be noted that all the elementary coefficients that are normalized per unit area have been derived and optimized with the surface-area calculation algorithms described here. Different surface area algorithms are likely to yield slightly different numerical results.

11. Structure-Based Thermodynamic Analysis of SH3 Domain

The application of the structural parameterization to the global stability of a protein will be illustrated with the SH3 domain of α-spectrin. SH3 domains are small (about 60 amino-acid residues), highly conserved, ubiquitous protein modules that are found in many proteins involved in intracellular signal transduction ***(41–45)***. SH3 domains mediate protein-protein interactions and modulate enzyme activity through their ability to bind specific proline-rich peptides. The structure of several SH3 domains is known at high resolution. In general, SH3 domains fold as compact beta barrels with five to six anti-parallel β strands. They form a nonpolar grove which is complementary to peptides in a polyproline-II conformation. Binding to these peptides control the activity of the signalling proteins in which they are found. The structure of the SH3 domain of α-spectrin has been solved crystallographically at 1.8Å resolution (pdb file 1 shg) ***(46)***. Furthermore, the thermodynamics of the SH3 domain from α-spectrin has been measured by high-sensitivity differential scanning calorimetry and other techniques ***(44,47)***. The results for the SH3 domain illustrate the need to explicitly consider the atomic packing density in enthalpy calculations.

11.1 The Heat Capacity Change

As discussed above, the primary contributions to the heat capacity change upon protein unfolding arise as a result of the exposure to solvent of nonpolar and polar groups that were buried in the native state. For the SH3 domain of α-spectrin 3074 $Å^2$ of nonpolar, 1922 $Å^2$ of polar and 110 $Å^2$ of hydroxy surface area become exposed to the solvent upon unfolding. These changes in accessibility yield a ΔC_p of 902 cal/K·mol. The experimental values range between 814 and 933 cal/K·mol ***(44,47)*** which bracket the value calculated from structure. Additional contributions such as those associated with the protonation of carboxylic groups are estimated to contribute very little (about 36 cal/K·mol).

11.2. The Enthalpy Change

At 60°C, the enthalpy change for unfolding of the SH3 domain predicted by the structural parameterization is 44.5 kcal/mol, which compares very well with the value of 43 kcal/mol obtained experimentally in H_2O and the value of 45 kcal/mol obtained in D_2O ***(44,47)***. However, it must be noted that a simple calculation based upon the change in surface accessibility alone will yield the wrong answer. Using the simple ΔASA scaling, a ΔH ***(60)*** of 34.4 kcal/mol will be obtained which underestimates significantly the experimental value. The difference arises primarily from a higher-than-average atomic packing density of those atoms that are buried within the structure of the SH3 domain. For the SH3 domain of α-spectrin, the average number of noncovalently bonded atoms within a distance of 3 Å is 4.6 compared to an average number of 4.4–4.5 for other proteins. Because of the atomic composition of this protein, these closely packed atoms tend to be less polar than average, increasing the relative contributions of nonpolar van der Waals interactions to the enthalpy change.

11.3. The Entropy Change

The entropy change upon unfolding consists mainly of two opposing contributions—the conformational entropy increase of backbone and side chains, and the negative change caused by the solvation of groups that are buried from the solvent in the native state. The conformational entropy change amounts to 323 cal/K·mol and the solvation entropy to –295 cal/K·mol, yielding a combined total of 28 cal/K·mol at 25°C. Together with the enthalpy change, the structural parameterization predicts a Gibbs energy of stabilization of 4.6 kcal/mol at 25°C in excellent agreement with the experimental value of 4.5 kcal/mol obtained from DSC ***(44,47)***. This example demonstrates that the solvation and conformational contributions to the entropy are of opposite sign, and about one order of magnitude larger than the total value. Since the solvation entropy scales in terms of ΔASA_i, most errors are usually compensated with errors of the opposite sign on the calculated enthalpy, which also scales to a large extent in terms of ΔASA_i. As a result, the calculated conformational entropy change is usually the most important source of error in the calculation of the Gibbs energy because it is not compensated by other changes.

11.4. Evaluation of Overall Energetics

The results of **Subheading 11.** indicate that changes in surface accessibility account well for heat-capacity changes—reinforcing the theory that these changes originate from local interactions with the solvent ***(32,33)***. Calculation of enthalpy changes, on the other hand, require an accurate estimation of van der Waals interactions in the native state. This estimation must be done in a way that avoids the subtraction of large numbers inherent to separate calculations of internal interactions and hydration. In our calculations, we decided to make the coefficients in **Eq. 4** a function of the interior packing in the protein. Thus far, the best agreement with the experimental results have been obtained when the the coefficients are scaled in terms of the average number of nonbonded atoms that are within 3 Å of solvent-buried atoms within the protein.

These above results indicate that the structural parameterization of the energetics accounts accurately for the energetics of stabilization of the SH3 domain of α-spectrin, and that the partitioning of the energetics into its enthalpic and entropic components is captured correctly. This comparison validates the use of this approach in the analysis of this protein and in the design of specific mutants.

12. Stability Mutants

Once the wild-type protein has been characterized, the usual task in the design of stability is the selection of appropriate locations for amino-acid mutations. The goal here is to increase ΔG, which can be done by increasing ΔH, decreasing ΔS, or an appropriate combination of both. One of the most pervasive problems in stability design is that favorable changes in enthalpy are usually coupled to unfavorable changes in entropy and vice versa—a phenomenon known as enthalpy-entropy compensation ***(48)***. In addition, the magnitude of each term is the differential between internal interactions and interactions with the solvent. For those reasons, the effects of mutations are context-dependent, and must be considered globally. The total Gibbs energy of a protein cannot be decomposed as a sum of individual contributions that can be calculated a priori in a generic fashion. There are many examples of stability mutations in the literature, including the analysis of thermophilic proteins. In the following paragraphs (*see* **Subheadings 12.1.** and **12.2.**), we discuss two situations. The reader is referred to other references for comprehensive articles or reviews ***(49–55)***.

12.1. Surface Mutations

From a structural point of view, mutations can be classified as surface mutations and internal mutations. Surface mutations are mutations at solvent-exposed locations, whereas internal mutations are those at locations that are buried from the solvent. Effects caused by surface mutations are easier to predict from the wild-type structure because they usually do not trigger a repacking of the protein core. Of all stabilizing mutations, Gly->Ala mutations at completely exposed surface locations are particularly attractive because their effect is primarily the result of a decrease in the conformational entropy change upon unfolding. Since they are solvent-exposed, enthalpy and solvent entropy differences between unfolded and native states are usually small. This type of mutation is one of the few—if not the only one—in which contextual effects are minimal. The presence of the β carbon in the Ala mutant reduces the number of degrees of freedom in the unfolded state, thus lowering its conformational entropy and therefore the conformational entropy gain associated with unfolding. The effect of this mutation is primarily on the unfolded rather than the native state. The effects of Gly->Ala mutations at surface-exposed locations on the Gibbs energy of stabilization have been investigated in different proteins ***(16,53,56)***. The experimental values range between 0.9 kcal/mol for RNase T1 ***(53)***, 0.96 kcal/mol for T4 lysozyme ***(56)***, 0.91 kcal/mol for barnase ***(57)***, 0.71 kcal/mol for a fragment (λ_{6-85}) of lambda repressor ***(58)***, and 0.73 kcal/mol for GCN4 ***(16)***. These values are close to each other, and also close to the expected values arising from backbone conformational entropy differences alone ***(16)***. Other surface mutations involve significant changes in solvent accessibility, and therefore their effects are context-dependent, including enthalpic and solvation entropy contributions in addition to backbone and side-chain conformational entropy differences. Accordingly, the same mutation can yield different results depending on the specific environment. For this reason, these effects need to be computed on a case-by-case basis.

12.2. Internal Mutations

Internal mutations are inherently more complex because they are invariable associated with a repacking of the protein core. Recently, Malakauskas and Mayo ***(59)*** designed a thermostable version of the Streptococcal protein G beta 1 domain (Gbeta1) by introducing eight mutations, including core mutations (V1M, Y4F, V7I, T17I, T19I, T26E, V30I, and V40I). Since the structures of both versions are known, we performed a structure-based thermodynamic analysis to explore the reasons for the enhanced stability of the mutant protein. The structural parameterization correctly predicted the higher stability of the thermostable version. The calculated Gibbs energy was 13.2 kcal/mol at 25°C, and the denaturation temperature was 103°C, in good agreement with the experimental values. According to the structural parameterization of the energetics, the reasons for enhanced stability cannot be attributed to a single factor. The heat-capacity change of the thermostable mutant is predicted to be slightly larger (1 kcal/K·mol vs 0.85 kcal/K·mol), consistent with a more hydrophobic core. The enthalpy change at the reference temperature is also larger (62.6 kcal/mol vs 54 kcal/mol), reflecting better van der Waals contacts in the mutant protein.

13. Protein Stability and Function

During the course of their biological function, proteins interact with other macromolecules and with small ligands. As a result of these interactions, the properties of the proteins can be modified in different ways. Interaction with ligands or other molecules can switch proteins from one functional state to another by selectively stabilizing some conformations. An example is shown in **Fig. 4** for a hypothetical enzyme and a noncompetitive allosteric inhibitor. The enzyme in the figure can exist in an active and an inactive conformation. In the absence of the allosteric inhibitor, the Gibbs energy of the active conformation is lower than that of the inactive conformation, so that the only form that is significantly populated is the active one. It must be understood that for a conformation to account for 99% of the population, a Gibbs energy difference of only 2.7 kcal/mol is required. The inactive form is able to bind the inhibitor with the affinity constant K_a. Accordingly, in the presence of the inhibitor, the Gibbs energy of the inactive form will be modified by an amount equal to $-RT\ln(1 + K_a[I])$ (for a general discussion, *see* **ref. *60***). An increase in the inhibitor concentration will stabilize the inactive conformation by lowering its Gibbs energy. For example, when the concentration of inhibitor is such that the product $K_a[I] = 1000$, the Gibbs energy of the inactive conformation will be favored by an additional –4 kcal/mol. This amount may be enough to make the inactive conformation the predominantly populated con-

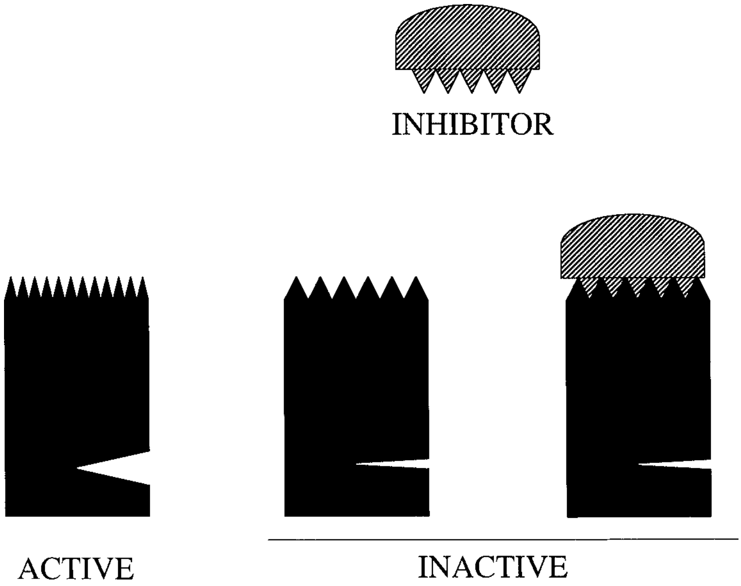


Fig. 4. A hypothetical enzyme and an allosteric inhibitor. The enzyme can exist in an active and an inactive conformation. In the absence of the inhibitor, the Gibbs energy of the active conformation is lower than that of the inactive conformation, that the only form that is significantly populated is the active one. Only the inactive form binds the inhibitor. An increase in the inhibitor concentration will stabilize the inactive conformation by lowering its Gibbs energy. Because the inhibitor binding site is distal from the active site, a thermodynamic coupling exists between these two sites: i.e., when the protein region where the inhibitor site is located is in a binding-competent conformation, the probability that the active site is in the closed conformation is very high. Conversely, if the inhibitor binding site is not in a binding-competent conformation, the probability that the active site is in the closed conformation is very low. This coupling in the behavior of two distal regions throughout the protein structure defines its functional cooperativity, and can be described by a set of joint probabilities ($P[i|j]$).

formation, thus switching the enzyme system from active to inactive. This example illustrates that strategies aimed at reengineering proteins for increased stability must take into consideration the effects of mutations on the relative population of different conformations, as well as on the binding affinity of ligands.

14. Functional Cooperativity and Protein Stability

In the example in **Fig. 4**, the interaction between inhibitor and protein occurs at a specific site (the inhibitor binding site) but, the resulting effect is detected at distal regions from the origin (the active site). For this phenomenon to occur, there must be a functional coupling between the inhibitor binding site and the active site—i.e., when the protein region where the inhibitor site is located is in a binding-competent conformation, the probability that the active site is in the closed conformation is very high. Conversely, if the inhibitor binding site is not in a binding-competent conformation, the probability that the active site is in the closed conformation is very low. This coupling in the behavior of two distal regions throughout the protein structure will be referred to as functional cooperativity, and can be described by a set of joint probabilities ($P[i|j]$). If the inhibitor and active sites are independent of each other, ligand-induced regulation is impossible. The functional cooperativity of a protein can be defined in terms of the structural distribution of atomic interactions within the protein. This distribution determines which regions of the protein are highly stable in a particular conformation, and which regions have a high probability of undergoing conformational excursions from their energy minima. The ability to predict the propagation of binding effects from one region to another has been demonstrated recently for the binding of the monoclonal antibody D1.3 to lysozyme. ***(61)***. The crystallographic structure for the complex between those two proteins has been determined at high resolution ***(62)***, and hydrogen-exchange data for both free lysozyme and its complex with D1.3 are available ***(63)***.

Williams et al. ***(63)*** observed that when D1.3 binds to lysozyme, the amino acids located at the binding interface acquire additional protection to hydrogen/deuterium exchange, as well as a number of residues distal from the binding interface. These distal residues showed enhanced protection because of an increase in their structural stability resulting from the binding of the antibody. The important point is that not every residue acquired extra protection—only a well-defined subset. From the point of view of functional analysis and design, the question is: Can this subset be predicted from structure? In order to answer this question, we used the following strategy:

1. Identify the structural determinants (hot spots) within the binding interface.
2. Using the high-resolution structure as a template, generate a large ensemble of conformations in which each conformation is characterized by having some region(s) unfolded and the remaining region(s) in the native conformation.
3. Estimate the Gibbs energy of each of the conformations generated with the computer

$$(\Delta G_i = \Delta G_i^0 - R\,T \ln \frac{(1 + K_{a,i}[x])}{(1 + K_{a,0}[x])}).$$

4. Those states in which the binding hot spots are formed will have their Gibbs energy reduced by an amount equal to $-RT\ln(1 + K_{a,0}[x])$ in relation to those states in which the binding site is not structured. Therefore, their probability will increase relative to those states in which the binding site is not formed.
5. Calculate the stability constants per residue for lysozyme in the absence and in the presence of the antibody. The implementation of this procedure is the Core_Bind algorithm ***(64)***.

The stability constants per residue, κ_{fj}, are the quantities that one would measure if it were possible to experimentally determine the stability of the protein by monitoring each individual residue. They are defined as the ratio of the summed probabilities of all states in which a residue is folded, to the summed probabilities of all states in which that residue is not folded ***(65)***:

$$\kappa_{f,j} = \frac{\sum_i P_{f,j,i}}{\sum_i P_{nf,j,i}} \tag{10}$$

For a significant number of residues in a protein, the hydrogen-exchange protection factors provide good estimates of κ_{fj} ***(47,65–68)***. The apparent Gibbs energy per residue is given by $\Delta G_{fj} = -RT\ln\kappa_{fj}$. If a residue only becomes exposed to solvent as a result of global unfolding, ΔG_{fj} is equal to the global Gibbs energy (ΔG_U). If local unfolding processes are also able to expose a residue to solvent, then $\Delta G_{fj} < \Delta G_U$.

Figure 5 illustrates the calculated effects of D1.3 binding on the residue stability constants for HEWL. In this figure, the residues in direct contact with the antibody are shown in black at the top. These residues are expected to show enhanced protection to exchange because of thermodynamic (increased stability) as well as steric (burial from solvent as a result of direct interaction with D1.3) reasons. Residues that are not in contact with the antibody, but are predicted by the Core_Bind algorithm to show enhanced protection because of the redistribution of the native state ensemble, are shown in light grey and highlighted by arrows. Finally, shown in dark are those residues not expected to exhibit significant changes in their protection factors. As shown in **Fig. 5**, the effects of D1.3 binding extend through most of the stable core of the HEWL molecule. The effect propagates from the origin (F helix) to the immediately adjacent B helix (residues 24–37) and neighboring residues, parts of the A helix (residues 8–15) and the loop connecting them, parts of the D helix (residues 108–115), the central loop of the β-domain (residues 55–65) and the adjacent strand (residues 50–54). These results are in excellent agreement with the experimental protection factors measured for the D1.3-HEWL complex (*see* **Fig. 3** and **Table 1** in **ref.** ***63***). There were 23 residues remote from the binding site for which enhanced protection could be measured. These residues are distributed throughout the regions predicted to show enhanced protection, except for two residues in helix C (Cys94 and Lys97), two residues (Cys76, Ile78) in the loop connecting the β-domain and helix E, and one residue (Leu84) in helix E, which fall outside the regions predicted to show enhanced protection. The arrows in **Fig. 5** define the path of cooperative interactions initiated by D1.3 binding. This path extends to regions remote from the binding site and to a significant portion of the HEWL molecule, despite the fact that the main binding determinant is one of the least stable parts of the molecule.

15. Thermodynamic Characteristics of Binding Sites

A significant portion of the binding site for the D1.3 antibody is defined by one of the least stable parts of the HEWL molecule (the

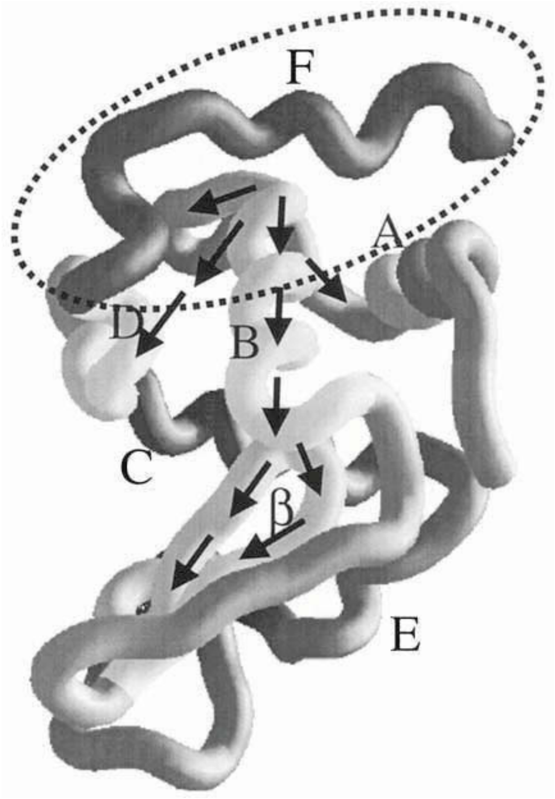


Fig. 5. The structure of HEWL showing the calculated effects of D1.3 binding on the stability constants per residue. In this figure, the residues in direct contact with the antibody are shown in black within a circle at the top. These residues are expected to show enhanced protection to exchange caused by thermodynamic (increased stability) as well as steric (burial from solvent as a result of direct interaction with D1.3) reasons. Shown in light grey and highlighted by arrows are residues that are not in contact with the antibody, but are predicted by the Core_Bind algorithm to show enhanced protection because of the redistribution of the native state ensemble. Finally, shown in dark are those residues not expected to exhibit significant changes in their protection factors. The helices and β domain have been labeled on the structure. The stability constants were calculated with the Core_Bind algorithm as described in ***(61)***, and the figure was made using the program GRASP ***(71)***.

F helix), and this characteristic is responsible for the initiation of the propagation of the effects originating at the binding site. We have also studied the distribution of residue stabilities in the binding site of the HIV-1 protease, and found that it had a dual character defined simultaneously by regions of high and low stability ***(64,69,70)***. Examination of other proteins reveals that this may be a common structural characteristic of binding sites ***(61)***. As in the case of the binding of the D1.3 antibody to HEWL, ligand binding will cause a redistribution in their native-state ensemble and a concomitant change in the residue stability constants of residues distal to the binding site. The magnitude of these effects is related to the stability of the residues defining the binding site. If the binding site has been formed by high-stability residues only, all states in the ensemble will be binding-competent and ligand binding would only induce an energy shift without an internal reordering in the ensemble. In this case, the propagation of the effects would be limited to the immediate neighborhood of the binding site. A ligand-induced redistribution in the probabilities of conformational states requires that only a subset of the native ensemble is binding-competent. This condition is satisfied when part of the residues that define the binding site exhibit low structural stability or exist in a nonbinding-competent conformation in the unligated protein.

16. Conclusion

Protein stability and functional cooperativity on allosteric regulation are linked functions related to each other by the distribution of atomic interactions throughout the protein structure. For that reason, protein engineering protocols aimed at creating stability mutants that retain full functionality must consider the potential effects of mutations on the distribution of cooperative interactions. The examples presented in this chapter suggest possible ways to address these issues. Supported by grants from the National Institutes of Health GM 51362 and GM 57144 and the National Science Foundation (MCB-9816661).

References

1. Murphy, K. P. and Gill, S. J. (1991) Solid model compounds and the thermodynamics of protein unfolding. *J. Mol. Biol.* **222,** 699–709.
2. Freire, E. and Murphy, K. P. (1991) The molecular basis of cooperativity in protein folding. *J. Mol. Biol.* **222,** 687–698.
3. Livingstone, J. R., Spolar, R. S., and Record, M. T., Jr. (1991) Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry* **30,** 4237–4244.
4. Murphy, K. P. and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* **43,** 313–361.
5. Straume, M., Murphy, K. P., and Freire, E. (1992) Thermodynamic strategies for protein design: increased temperature stability, in *Biocatalysis at Extreme Temperatures: Enzyme Systems Near and Above 100 °C.* (Adams, M. W. W. and Kelly, R. M., eds.), ACS Books, Washington, D.C., pp. 122–135.
6. Murphy, K. P., Bhakuni, V., Xie, D., and Freire, E. (1992) The molecular basis of co-operativity in protein folding III: identification of cooperative folding units and folding intermediates. *J. Mol. Biol.* **227,** 293–306.
7. Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **31,** 3947–3955.
8. Murphy, K. P., Xie, D., Garcia, K. C., Amzel, L. M., and Freire, E. (1993) Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins Struct. Funct. Genet.* **15,** 113–120.
9. Murphy, K. P., Xie, D., Thompson, K., Amzel, L. M., and Freire, E. (1994) Entropy in biological binding processes: estimation of translational entropy loss. *Proteins Struct. Funct. Genet.* **18,** 63–67.
10. Xie, D. and Freire, E. (1994) Structure based prediction of protein folding intermediates. *J. Mol. Biol.* **242,** 62–80.
11. Xie, D. and Freire, E. (1994) Molecular basis of cooperativity in protein folding. V. thermodynamic and structural conditions for the stabilization of compact denatured states. *Proteins Struct. Funct. Genet.* **19,** 291–301.
12. Spolar, R. S. and Record, M. T. (1994) Coupling of local folding to site specific binding of proteins to DNA. *Science* **263,** 777–784.

13. Freire, E. (1993) Structural thermodynamics: prediction of protein stability and protein binding affinities. *Arch. Biochem. Biophys.* **303** 181–184.
14. Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994) Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins Struct. Funct. Genet.* **20,** 68–84.
15. Gomez, J., Hilser, J. V., Xie, D., and Freire. E., (1995) The heat capacity of proteins. *Proteins Struct. Funct Genet.* **22,** 404–412.
16. D'Aquino, J. A., Gómez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25,** 143–156.
17. Hilser, V. J., Gomez, J., and Freire, E. (1996) The enthalpy change in protein folding and binding. Refinement of parameters for structure based calculations. *Proteins* **26,** 123–133.
18. Luque, I., Mayorga, O., and Freire, E. (1996) Structure based thermodynamic scale of α-helix propensities in amino acids. *Biochemistry* **35,** 13,681–13,688.
19. Bardi, J. S., Luque, I., and Freire, E. (1997) Structure-based thermodynamic analysis of HIV-1 protease inhibitors. *Biochemistry* **36,** 6588–6596.
20. Luque, I., Gomez, J., Semo, N., and Freire, E. (1998) Structure-based thermodynamic design of peptide ligands. Application to peptide inhibitors of the aspartic protease endothiapepsin. *Proteins* **30,** 74–85.
21. Robertson, A. D. and Murphy, K. P. (1997). Protein structure and the energetics of protein stability. *Chem. Rev.* **97,** 1251–1267.
22. Brandts, J. F. (1964) The thermodynamics of protein denaturation. I. The denaturation of chymotrypsin. *J. Am. Chem. Soc.* **86,** 4291–4301.
23. Lumry, R., Biltonen, R., and Brandts, J. F. (1966) Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* **4,** 917–944.
24. Lumry, R. and Biltonen, R. (1969) Thermodynamic and kinetic aspects of protein conformations in relation to physiological function, in *Biological Macromolecules: Structure and Stability of Biological Macromolecules.* Vol. II. S. Timasheff and G. Fasman, eds., Marcel Dekker, New York, N.Y, pp. 65–211.
25. Thompson, K., Vinson, C., and Freire, E. (1993) Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. *Biochemistry* **32,** 5491–5496.

26. Johnson, C. R., Morin, P. E., Arrowsmith, C. H., and Freire, E. (1995) Thermodynamic analysis of the structural stability of the tetrameric oligomerization domain of p53 tumor suppresor. *Biochemistry* **34,** 5309–5316.
27. Johnson, C. R. and Freire, E. (1996) Structural stability of oligomeric proteins. *Techniques in Prot. Chem.* **VII,** 459–467.
28. Freire, E. (1995) Differential scanning calorimetry, in *Protein Stability and Folding,* vol. 40. (Shirley, B., ed.), pp. 191–218.
29. Freire, E. (1995) Thermal denaturation methods in the study of protein folding. *Methods Enzymol.* **259,** 144–168.
30. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Computational Chem.* **4,** John Wiley, New York, NY, pp. 187–217.
31. Lazaridis, T., Archontis, G., and Karplus, M. (1995) Enthalpic contributions to protein stability: atom-based calculations and statistical mechanics. *Adv. Prot. Chem.* **47,** 231–306.
32. Suurkuusk, J. (1974) Specific heat measurements on lysozyme, chymotrypsinogen and ovalbumin in aqueous solution and in the solid state. *Acta Chem. Scand. B.* **28,** 409–417.
33. Yang, P. H. and Rupley, J. A. (1979) Protein-water interactions. Heat capacity of the lysozyme-water system. *Biochemistry* **18,** 2654–2661.
34. Gomez, J. and Freire, E. (1995) Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *J. Mol. Biol.* **252,** 337–350.
35. Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. USA* **83,** 8069–8072.
36. Cabani, S., Gianni, P., Mollica, V., and Lepori, L. (1981) Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Solution Chem.* **10,** 563–595.
37. García-Moreno, B. E. (1995) Probing structural and physical basis of protein energetics linked to protons and salt. *Methods Enzymol.* **259,** 512–538.
38. García-Moreno, B. E., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S., and Stites, W. E. (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **4,** 211–224.
39. Lee, B. and Richards, F. M. (1971) The Interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400.

40. Creamer, T. P., Srinivasan, R., and Rose, G. D. (1995) Modeling unfolded states of peptides and proteins. *Biochemistry* **34,** 16,245–16,250.
41. Kuriyan, J. and Cowburn, D. (1993) Structures of SH2 and SH3 domains. *Curr. Opin. Struct. Biol.* **3,** 828–837.
42. Morton, C. J. and Campbell, I. D. (1994) Molecular "Velcro." *Curr. Biol.* **4,** 615–617.
43. Pawson, T. (1995) Protein modules and signalling networks. *Nature* **373,** 573–580.
44. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994) Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry* **33,** 2142–2150.
45. Xu, W., Harrison, S. C., and Eck, M. J. (1997) Three dimensional structure of the tyrosine kinase c-Src. *Nature* **385,** 595–602.
46. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992) Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359,** 851–855.
47. Sadqi, M., Casares, S., Abril, M. A., Mayorga, O. L., Conejero-Lara, F., and Freire, E. (1999) The native state conformational ensemble of the SH3 domain from α-spectrin. *Biochemistry* **38,** 8899–8906.
48. Lumry, R. and Rajender, S. (1970) Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers* **9,** 1125–1227.
49. Gilis, D. and Rooman, M. (1996) Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. *J. Mol. Biol.* **257,** 1112–1126.
50. Gilis, D. and Rooman, M. (1997) Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. *J. Mol. Biol.* **272,** 276–290.
51. Vogt, G., Woell, S., and Argos, P. (1997) Protein thermal stability, hydrogen bonds and ion pairs. *J. Mol. Biol.* **269,** 631–643.
52. Myers, J. K., Smith, J. S., Pace, C. N., and Scholtz, J. M. (1996) The α-helix of ribonuclease T1 as an independent stability unit: direct comparison of peptide and protein stability. *J. Mol. Biol.* **263,** 390–395.
53. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1997) Helix propensities are identical in proteins and peptides. *Biochemistry* **36,** 10,923–10,929.

54. Takano, K., Yamagata, Y., and Yutani, K. (1998) A general rule for the relationship between hydrophobic effect and conformational stability of a protein: stability and structure of a series of hydrophobic mutants of human lysozyme. *J. Mol. Biol.* **280,** 749–761.
55. Hollien, J. and Marqusee, S. (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases h. *Biochemistry* **38,** 3831–3836.
56. Blaber, M., Zhang, X.-j., Lindstrom, J. L., Pepiot, S. D., Baase, W. A., and Matthews, B. W. (1994) Determination of alpha-helix propensity within the context of a folded protein: sites 44 and 131 in bacteriophage T4 lysozyme. *J. Mol. Biol.* **235,** 600–624.
57. Horovitz, A., Matthews, J. M., and Fersht, A. R. (1992) Alpha-helix stability in proteins. II. factors that influence stability at an internal position. *J. Mol. Biol.* **227,** 560–568.
58. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by Ala*Gly substitutions. *Nat. Struct. Biol.* **4,** 305–310.
59. Malakauskas, S. M. and Mayo, S. L. (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5,** 470–476.
60. Freire, E. (1997) The statistical thermodynamic linkage between conformational and binding equilibrium. *Adv. Protein Chem.* **51,** 255–279.
61. Freire, E. (1999) The propagation of binding interactions to remote sites in proteins. Analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl. Acad. Sci. USA* **96,** 10,118–10,122.
62. Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., et al. (1994) Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* **91,** 1089–1093.
63. Williams, D. C., Benjamin, D. C., Poljak, R. J., and Rule, G. S. (1996) Global changes in amide hydrogen exchange rates for a protein antigen in complex with three different antibodies. *J. Mol. Biol.* **257,** 866–876.
64. Todd, M. J. and Freire, E. (1999) The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **36,** 147–156.

65. Hilser, V. J. and Freire, E. (1996) Structure based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **262,** 756–772.
66. Hilser, V. J., Townsend, B. D., and Freire, E. (1997) Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: structural mapping of cooperative interactions. *Biophys. Chem.* **64,** 69–79.
67. Hilser, V. J. and Freire, E. (1997) Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease. *Proteins* **27,** 171–183.
68. Hilser, V. J., Oas, T., Dowdy, D., and Freire, E. (1998) Structural identification of key residues in protein folding: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. USA* **95,** 9903–9908.
69. Luque, I., Todd, M. J., Gomez, J., Semo, N., and Freire, E. (1998) The molecular basis of resistance to HIV-1 protease inhibition: a plausible hypothesis. *Biochemistry* **37,** 5791–5797.
70. Todd, M. J., Semo, N., and Freire, E. (1998) The structural stability of the HIV-1 protease. *J. Mol. Biol.* **283,** 475–488.
71. Nicholls, A., Bharadwaj, R., and Honig, B. (1993) GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **64,** 166–170.

4

Measuring the Conformational Stability of a Protein by Hydrogen Exchange

Beatrice M. P. Huyghues-Despointes, C. Nick Pace, S. Walter Englander, and J. Martin Scholtz

1. Introduction

Measuring the conformational stability of a protein is one key to solving the protein folding problem. It is also of practical importance for answering questions such as these:

1. How stable is a protein under physiological conditions?
2. How does the stability depend on temperature, pH, and salt concentration?
3. Can the stability be increased by osmolytes?
4. Can the stability be increased by ligands that bind the native state?
5. Does an amino-acid substitution increase or decrease the stability?

Thus, reliable techniques for measuring the conformational stability of proteins are essential. In this chapter, we briefly review the traditional methods that have been used to measure the conformational stability of a protein, and describe in more detail how the conformational stability can be measured using hydrogen-exchange rates as monitored by nuclear magnetic resonance (NMR).

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy

2. Determining Protein Conformational Stability

The conformational stability is defined as the free-energy change of the unfolding reaction, F $\Leftrightarrow$ U, under ambient conditions (generally around room temperature and neutral pH): $\Delta G_U = G_U - G_F$. It can be evaluated using the relationship $\Delta G_U = -RT \ln K_U$, where K_U is the equilibrium constant of the unfolding reaction. Measuring the conformational stability and changes in stability of proteins that differ slightly in structure helps to define the forces that determine the conformations of proteins and optimize their stabilities. The stability of the folded conformation is a delicate balance between compensating forces: favorable hydrophobic, van der Waals, and hydrogen-bonding interactions and unfavorable conformational entropy. The result is that the folded protein is only marginally stable, generally 5–15 kcal/mol under native conditions ***(1)***.

A protein exists predominantly in the folded conformation under ambient conditions, but rarely, as dictated by the Boltzmann relationship, $P_f/F = e^{-\Delta G_U/RT}$, this conformation will unfold to partially unfolded or completely unfolded higher energy conformations (*see* **Fig. 1**). As a result, it is difficult to detect the unfolded conformations and to directly measure K_U under native conditions. In the traditional methods, the conformational stability is extrapolated from stabilities obtained under denatured conditions, while in the hydrogen-exchange experiments, conformational stability is determined directly under native conditions.

2.1. Traditional Methods: Solvent and Thermal Denaturation

The most common way to estimate the conformational stability of a protein is to tilt the equilibrium to conditions where the folded and unfolded states are almost equally populated, measure the stability, and extrapolate to obtain the stability under native conditions. Traditional methods of measuring ΔG_U are solvent (urea or guanidinium chloride [GdmCl] or thermal denaturation (*see* **Fig. 2**) ***(2)***. Solvent denaturation curves are generally analyzed using the linear extrapolation method (LEM):

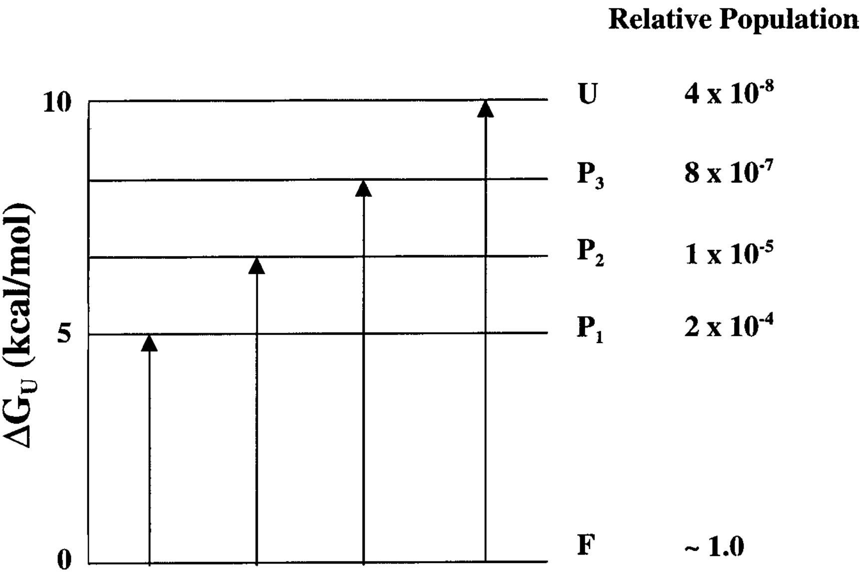


Fig. 1. An energy manifold of states at 25°C for a protein having ΔG_U = 10 kcal/mol, including the folded (F) and unfolded (U), and partially unfolded states (P_i).

$$\Delta G = \Delta G_U(H_2O) - m[\text{denaturant}] \tag{1}$$

where m is a measure of the dependence of ΔG on denaturant, and $\Delta G_U(H_2O)$ is an estimate of the conformational stability that assumes that the linear dependence of ΔG on denaturant observed in the transition region continues to 0 M denaturant. Thermal denaturation experiments yield the melting temperature, T_m, the enthalpy change at T_m, ΔH_m, and the heat-capacity change, ΔC_p, which can then be used to calculate ΔG_U at any temperature T, $\Delta G_U(T)$, with the Gibbs-Helmholtz equation:

$$\Delta G_U(T) = \Delta H_m(1 - T/T_m) + \Delta C_p[T - T_m - T \ln (T/T_m)] \tag{2}$$

The equilibrium between the folded and unfolded states can be monitored by calorimetry ***(3)***, or by using spectroscopic techniques such as UV absorbance spectroscopy, fluorescence, and circular dichroism that monitor the changes in conformational states of a protein ***(4)***.

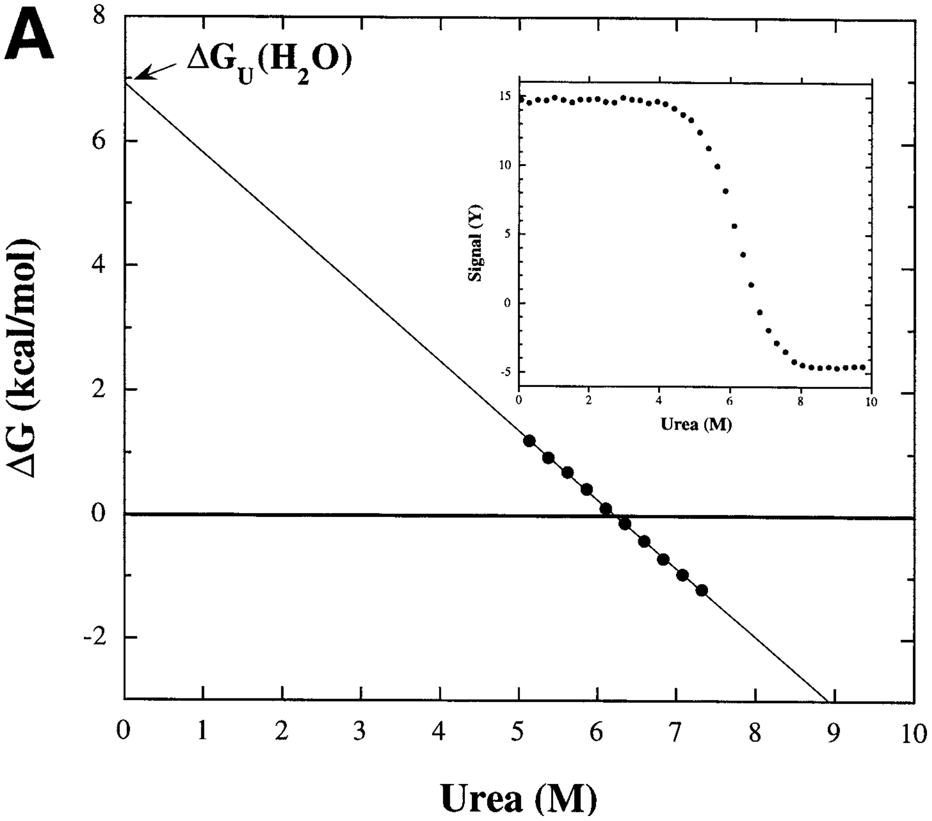


Fig. 2. **(A)** The inset shows a typical urea unfolding curve for a small monomeric protein. The data in the transition region can be extrapolated to determine $\Delta G_U(H_2O)$ using the LEM (**Eq. 1**).

There are several problems with these techniques. Traditional methods are usually unable to detect unfolded forms under ambient conditions because their concentrations are far too low. These methods for determining the conformational stability also depend on some basic limiting assumptions. First, they require that the folding transition of a protein be fully reversible, with a known and finite number of observable states. Above, we show the simplest example—a two-state mechanism—but the models can be expanded for analysis of multimeric protein systems or proteins that fold by mechanisms with intermediate states *(5)*.

B

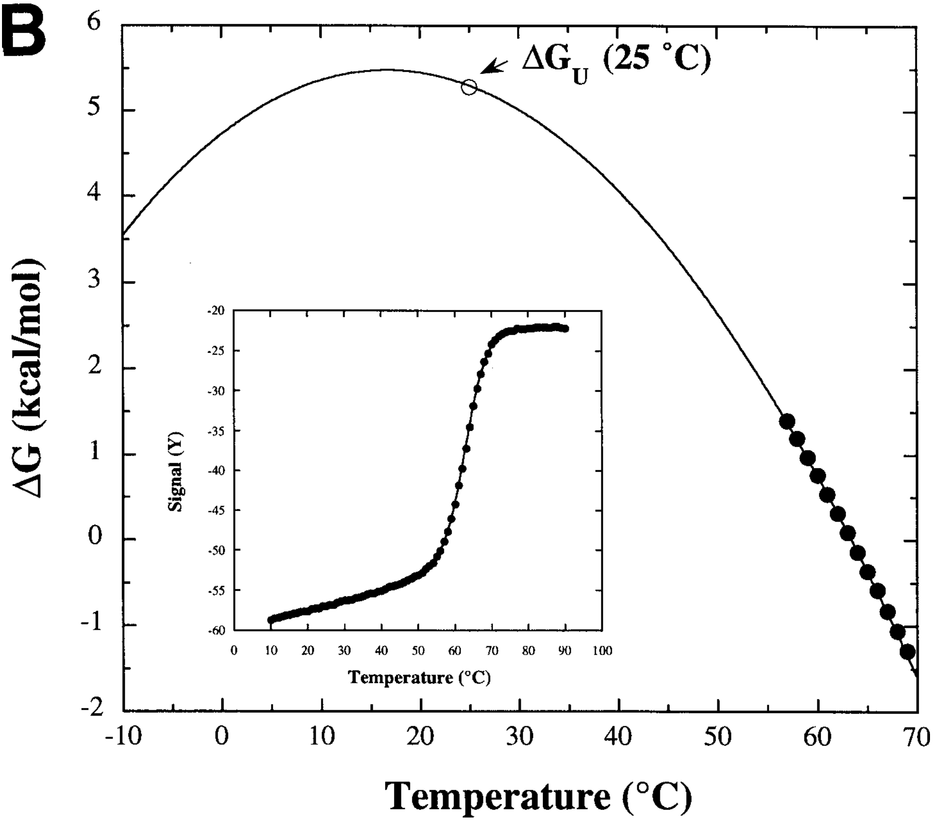


Fig. 2. *(continued)* **(B)** The inset shows a typical thermal unfolding curve for a small mononmeric protein. The data in the transition region of a thermal unfolding curve can be extrapolated to obtain ΔG_U at a desired reference temperature such as 25°C using the Gibbs-Helmholtz relationship (**Eq. 2**).

Second, both methods require a long extrapolation from unfolding to ambient conditions. The LEM assumes that the dependence on denaturant can be extended linearly to the absence of denaturant. This may not always occur, especially for GdmCl, a salt that can alter the stability of a protein at low denaturant concentrations through an ionic strength effect (*see* **ref.** ***6*** and references therein). For thermal denaturation, the extrapolation from higher temperatures to ambient temperatures requires an accurate value of ΔC_p. At best, ΔC_p can be measured to approx ± 10% with calorimetry.

Finally, the unfolded states in denaturant or at high temperatures may not be thermodynamically equivalent to the unfolded states under ambient conditions.

2.2. The New Method: Hydrogen Exchange

When a protein is placed in D_2O, amide hydrogens in the protein begin to exchange with deuterium. The rate of exchange of an amide depends on its environment in the protein. An unprotected amide exchanges readily with the solvent. In contrast, an amide that is shielded from solvent by burial and/or hydrogen bonding can exchange only a small fraction of time when it unfolds to a higher energy conformation that is exposed to solvent. The exchange of a backbone amide proton can occur by local fluctuations, such as the breaking of only a few hydrogen bonds, or by a larger global unfolding event. The result is that exchange rates of amide protons in a protein vary widely, providing a clear view of the higher energy unfolded states under native conditions (*see* **Fig. 1**) ***(7)***.

We describe here a NMR technique that identifies amides that exchange from the globally unfolded protein. Hydrogen-exchange measurements under fully native conditions can provide a direct estimate of the conformational stability of a protein without having to drive the protein through its macroscopic unfolding transition. This technique, which can measure the conformational stability of a protein, may not be possible by traditional methods because folding is irreversible or does not approach a two-or three-state folding mechanism.

3. The Hydrogen-Exchange Process

3.1. The Rate Constants Governing Exchange

The exchange process can be described using a general two-step model:

$$(\mathrm{N - H \cdots O = C})_{\mathrm{cl}} \underset{k_{\mathrm{cl}}}{\overset{k_{\mathrm{op}}}{\Leftrightarrow}} (\mathrm{N - H})_{\mathrm{op}} \xrightarrow{k_{\mathrm{rc}}} \mathrm{N - D} \tag{3}$$

where k_{op} and k_{cl} are the rate constants for structural opening and closing, respectively, and $(N-H)_{op}$ is the open form that exchanges with solvent at the intrinsic rate constant k_{rc}, where rc stands for random coil ***(8,9)***. The second step is written as an irreversible reaction because of exchange with a vast excess of solvent D_2O. The measurable rate constant $k_{ex} = k_{op}k_{rc}/(k_{op} + k_{cl} + k_{rc})$. If the protein is present mainly in the open form ($k_{op} >> k_{cl}$), $k_{ex} = k_{rc}$. If the protein is present mainly in the closed form ($k_{op} << k_{cl}$), then two mechanisms could potentially occur, depending on conditions, either as two distinct processes or in combination ***(10,11)***.

The EX1 exchange conditions (monomolecular exchange) exist when $k_{cl} << k_{rc}$, so that $k_{ex} = k_{op}$. The EX2 exchange conditions (bimolecular exchange) exist when $k_{cl} >> k_{rc}$, so that $k_{ex} = K_{op}k_{rc}$, where K_{op} is the equilibrium constant for structural opening (k_{op}/k_{cl}). Thus, under EX2 conditions, the free energy of structural opening is given by:

$$\Delta G_{HX} = -RT \ln (k_{ex}/k_{rc}). \quad (4)$$

One expects a subset of the ΔG_{HX} values to generally represent the complete unfolding reaction; thus, it has been proposed that the average of the three largest ΔG_{HX} values calculated using **Eq. 4** provides a reasonable estimate of the conformational stability of a protein ***(12)***.

3.2. The k_{rc} Values from Model Peptides

Exchange from the open form depends on k_{rc}. The k_{rc} value can be approximated by the chemical exchange rates of amide groups in nonstructured peptides. The rate constants in short (1–5 amino acids long) peptides and in poly-DL-alanine (PDLA) have been measured ***(9,13,14)***. The results show that exchange-rate constants of backbone amide protons are dependent on pH, temperature, identity of neighboring side-chain residues in the sequence, solvent isotope effects, and small ionic strength effects. The exchange-rate constants for alanine-based reference peptides at low ionic strengths and 20°C are shown in **Table 1**.

Table 1
H to D Exchange Rate Constants for Alanine-Based Models at 293 K[a]

Model	log k_A (M^{-1} min^{-1})	log k_B (M^{-1} min^{-1})	log k_W (min^{-1})
N-Ac-A-<u>A</u>-A-N′MA[b]	2.04	10.36	–1.5
PDLA[c]	1.62	10.05	–1.5

[a] These values are specific for NH to ND exchange under normal low salt conditions, using pD = pH* + 0.4 and pK_a = 15.05 for the D_2O at 20°C. These data are from **Table 3** in Bai et al. ***(9)***.

[b] L-alanine tripeptide blocked by N-acetyl at N-terminus and N′-methylamide at C-terminus.

[c] poly-DL-alanine.

The NH to ND chemical exchange process can be acid (k_H), base (k_{OH}), or water catalyzed (k_W), yet the latter generally makes only a negligible contribution:

$$K_{rc} = k_A\ 10^{-pD} + k_B\ 10^{[pD-pK_D]} + k_W \tag{5}$$

where pD is the corrected reading in D_2O (*see* below) and pK_D is the molar ionization constant of D_2O. The rate of exchange of a backbone amide reaches a minimum near pH ~ 3 and increases substantially at both pH extremes, by a factor of 10 per pD unit (**Eq. 5**). The chemical exchange rate of amide proton of residue *i* in a sequence is affected by its own side chain and the side chain of residue i–1. For example, we show in **Fig. 3** for two dipeptide sequences in RNase T1 that k_{rc} can vary by 20-fold. **Fig. 3** also illustrates the dependence of k_{rc} on pH.

The intrinsic rate of exchange can also change with temperature. The rate constant for the acid-catalyzed exchange can be modified by:

$$k_A(T) = k_A(293\ K)\ e^{(-Ea[1/T\ -\ 1/293]/R)} \tag{6}$$

where the activation energy, $Ea(k_H)$ = 14 kcal/mol. Analogous expressions can be made for the base- and water-catalyzed exchange in **Eq. 5**, where $Ea(k_{OH})$ = 17 kcal/mol, and $Ea(k_W)$ = 19 kcal/mol,

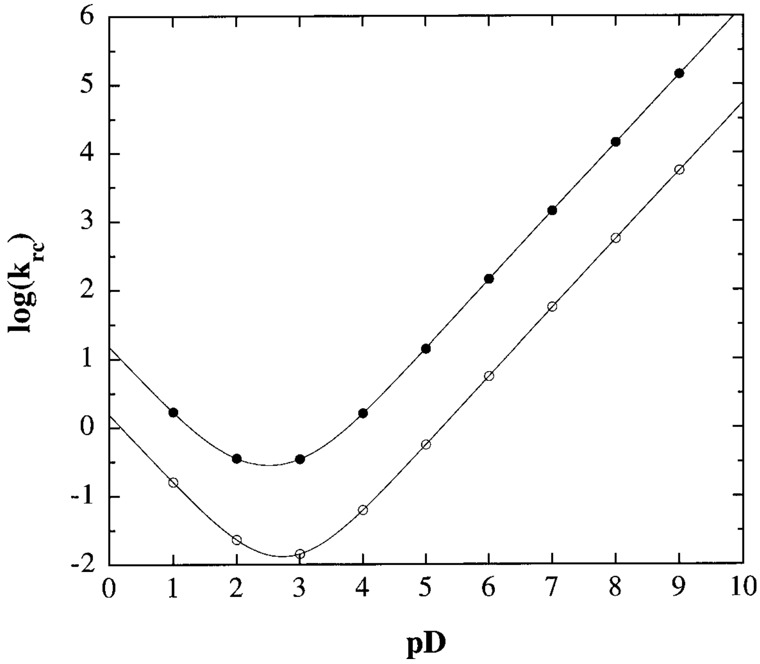


Fig. 3. Log (k_{rc}) vs pD for two dipeptide sequences, –Gly71-Ser72– (●) and –Val89-Ile90– (○), in RNase T1 using k_{rc} values from Bai et al. ***(9)***. Curves represent the data fit to **Eq. 5**.

respectively. To determine ΔG_{HX} in a protein, the exchange-rate constants of random coil peptides can be used to predict k_{rc} values for every residue at a particular pH and temperature. (*See* the spreadsheet at ftp://hxiris.med.upenn.edu/HX/ to do it for your favorite protein).

3.3. EX1 and EX2 Conditions

The exchange process for a globally exchanging residue must be EX2 in order to determine the conformational stability of the protein by the hydrogen-exchange method. Three methods are used to establish whether an amide proton exchanges by an EX1 or EX2 mechanism ***(15)***. The simplest method is to plot log k_{ex} for several amide protons at two different pH values above pH 4.0 and in a pH region where the stability of the protein changes as little as possible. For the EX2 process, the hydrogen-exchange rates for each NH group should change by 10-fold, with a one-unit change in pH, showing the expected pH dependence on k_{rc} (*see* **Fig. 3**). In the EX1 process, k_{ex} depends only on k_{op}, which is likely to be less dependent on pH. As a result, we would observe little to no rate variation with pH in the EX1 regime.

Another test of EX2 vs EX1 kinetics uses the nuclear Overhauser effect (NOE) ***(15–18)***. We can measure a decrease in the NOE intensity between two adjoining residues as hydrogen-exchange progresses. In the EX1 limit, two amide protons will exchange at the same rate, so that the decay rate of the NOE will be equal to the exchange rate of a single amide proton. Conversely, in the EX2 limit, two amide protons will exchange independently and often with different chemical exchange rates. Therefore, the decay rate of the NOE will be the sum of the exchange rates of the two adjoining amide protons. This method is often arduous, and errors in measurement can be significant ***(15)***.

A third method is simply to compare the ΔG_{HX} values for neighboring amide groups with their corresponding k_{rc} values. In an EX2 process, the same ΔG_{HX} values are likely to have widely different k_{rc} values. The opposite holds true for the EX1 process.

3.4. ΔG_{HX} for Measurable Amide Protons

The hydrogen-exchange technique provides a glimpse into the distribution of unfolded conformations at native conditions. Exchange of an amide proton with solvent deuterium occurs only through exchange-competent, higher-energy forms. The amide groups that are most protected from exchange have the largest stabilities and may be limited to exchange only from the fully unfolded state (“global”). Residues with intermediate and low stabilities (or protection) exchange by either fluctuations in the native state (“local”) or by partial unfolding (“subglobal”). Here, we describe an NMR technique that identifies only the global-exchanging residues. Monitoring hydrogen exchange in response to external perturbants, such as temperature, solvent denaturants ***(19–25)*** or pressure ***(26)***, can help to decipher the exchange processes of local and subglobal amide groups.

Hydrogen-exchange studies with mutant proteins have also been performed where the change in stability upon mutation, $\Delta\Delta G_{HX}$, can be obtained by:

$$\Delta\Delta G_{HX} = (\Delta G_{HX})_{WT} - (\Delta G_{HX})_{mutant} = -RT \ln(k_{ex})_{WT}/(k_{ex})_{mutant} \quad (7)$$

The k_{rc} values for all residues cancel in **Eq. 7**, except for the mutation site and the residue C-terminal to the site of the mutation. The residues that exchange by global unfolding will have $\Delta\Delta G_{HX}$ values similar to the values determined by the traditional methods. Unfortunately, there can also be nonglobal residues that have $\Delta\Delta G_{HX}$ values equal to or even greater than $\Delta\Delta G_U$ ***(27)***. Additional experiments that monitor the effect of an external perturbant on the hydrogen-exchange stabilities are required to decipher the stability patterns of nonglobal residues. As a result, hydrogen-exchange studies with mutant proteins cannot in themselves determine the residues that unfold globally.

4. The NMR Experiment

4.1. Establishing Experimental Conditions

The hydrogen-exchange experiment by NMR entails monitoring the decay of individual proton resonances over time as they exchange with solvent deuterium. We assume that standard sequential NMR assignment techniques have been used to identify amide resonances that persist upon the addition of deuterium. Normally, the overcrowding of proton resonances in a protein limits the usefulness of 1D proton spectra. Instead, popular two-dimensional experiments have been used, such as proton-detected magnitude-mode Correlation Spectroscopy (COSY), and more recently, inverse-detected ^{15}N Heteronuclear Single Quantum Coherrence (HSQC). For hydrogen exchange, the HSQC experiment provides better resolution when observing the exchange of many exchangeable backbone amide groups in larger proteins (15–25 kDa), and in most cases, HSQC spectra can be acquired in less time. The limitations to the heteronuclear approach are that the protein must be uniformly labeled with ^{15}N—usually by overproduction in bacteria, grown in minimal media where yields are generally reduced. ^{15}N-labeled precursors are expensive (although prices are steadily decreasing), and some degree of technical expertise of the NMR hardware and software is required.

A positive attribute of measuring the conformational stability by hydrogen exchange is that only the exchange rates of the most stable NH groups must be determined. This simplifies assignments and in most cases, simple 1D-proton spectra may be taken to collect the data. The most stable amide resonances are usually the ones that persist over the longest time. Most other resonances disappear from the spectra by the time the most stable amide protons begin to exchange. However, caution must be exercised, because the most stable residue may not always be the slowest exchanging residue. As shown in **Fig. 3** and **Eq. 4**, k_{rc} varies with sequence and can affect the exchange time and the stability value of a residue. To be safe, the exchange of at least the 10 slowest exchanging amide protons should be monitored.

It is important to establish the best pH and temperature conditions to monitor the decay of the most stable NH groups. The random coil rate of each residue and the stability of the protein play important roles in determining the best experimental conditions. For example, RNase T1 has a bell-shaped stability vs pH profile with a stability maximum of ~13 kcal/mol near pH 5.0 and 25°C in D_2O ***(28)***. If we monitor exchange at these conditions, the most stable residues would not exchange for over 100 years! As a result, experimental conditions must be chosen to allow the data to be collected in a reasonable amount of time. If we raise the pH or increase the temperature for RNase T1, the exchange is considerably faster (pD 7.7, 25°C—3–6 months, and 40°C—days). Another method to reduce the time of exchange at a desired experimental condition is to perform measurements at two or more concentrations of denaturant where the protein is still fully folded (the pretransition region of **Fig. 2A**), and extrapolate to the ΔG_{HX} in the absence of denaturant using **Eq. 1**. With both these techniques, it is important to test whether the exchange process is still in the EX2 regime. Conversely, if a protein is marginally stable, the temperature or the pH can be decreased in order to observe the decay of the most stable amide protons. Stabilities in a range of 4–10 kcal/mol can be measured using the hydrogen-exchange technique described here. A good starting point is to perform one-dimensional hydrogen-exchange experiments at several different conditions. If exchange is too slow or too fast, adjust the conditions until the most stable residues exchange in a reasonable amount of time and by the EX2 process.

4.2. Sample Preparation

NMR samples are prepared by dissolving the protein in an H_2O buffered solution. A 0.5–2 m*M* solution of protein (0.5–1.0 mL) is required. The one-dimensional experiment needs less protein than the two-dimensional experiment. The pH of this solution should be equal to the desired pD of the deuterated NMR sample. The pD of a solution in D_2O is the pH meter reading (pH*) plus 0.4 (pD = pH* + 0.4) ***(29)***. Measuring the pH* of a deuterated solution is performed as follows:

1. Use appropriate standard pH buffers in H_2O to calibrate your pH meter, considering the effect of temperature on the buffers.
2. Allow the electrode to soak in D_2O until the meter reading is steady (5–10 min).
3. Read the pH meter value of your sample (pH*) and make the appropriate correction for pD.

It is important to ensure that the protein sample is fully dissolved and at equilibrium before exchanging the H_2O for D_2O. If any particulate remains, centrifuge the sample and place the solution in a fresh Eppendorf tube. Buffers that are silent in the NMR experiment should be used—such as phosphate that lacks protons or one such as acetate, which can be purchased inexpensively in a deuterated form. The sample can be exchanged into D_2O by using a spin-column (Isolab, Inc.). The spin-column approach can be performed as follows:

1. Swell 1 g of G-25 resin in 10 mL of D_2O for a minimum of 3 h.
2. Remove fines by stirring and replacing with an equal vol of fresh liquid (2–3 times), pour deuterated buffer onto the resin for the last volume exchange, and equilibrate for 10 min.
3. Pour the resin into the spin column (~3 mL) and wash the resin several times with 0.5–1.0 mL volumes of deuterated buffer, by spinning the column in a swinging bucket benchtop centrifuge at ~1000*g* for 1-min intervals. Be sure to balance the rotor by using a dummy column with resin in it.
4. Exchange the protein solution in H_2O with D_2O by loading the sample in H_2O onto the column and spinning the column for a 1-min interval. The number of spins and the vol of sample added can be adjusted so that a final vol of recovered sample is ~ 0.7 mL and protein yield is 80–90%. (After the experiment, if the protein is valuable, one can rescue the rest of the protein left on the column by rinsing with several volumes of buffer.)
5. Immediately after exchange, transfer the solution to a NMR tube, and quickly adjust the pH* of your solution with NaOD or DCl if it has changed. The pH* in the NMR tube can be measured using a long, thin electrode from Ingold.
6. Place the sample in the NMR spectrometer with the probe temperature set at the appropriate temperature, and acquire a data set.
7. Measure the pH* of the sample after the conclusion of the NMR data set.

4.3. NMR Data and Analysis

NMR spectra must be collected over time. The NMR sample should be equilibrated in a water bath at the appropriate temperature if the NMR data are not acquired in a single time period. The spectra are processed by standard processing programs. The peak intensities for each residue can be measured either by peak heights (one-dimensional experiments) or by volume integration (two-dimensional crosspeaks). These measurements can fit to a single exponential function, $A = A_0[e^{(-k_{ex}t)}] + C$, where *C* is the baseline noise of a particular spectra caused by residual water in the sample. With partial or completely overlapping peaks, the exchange data may be fit to a sum of exponentials, especially if rates for amide protons differ by more than approximately threefold. The exchange rate should be measured only after the peak is approx 50% exchanged. The errors in stability are the largest for residues whose hydrogen exchange rates are either very fast or very slow. For a typical exchange experiment, two fold variations in the rate only affect ΔG_{HX} by 0.4 kcal/mol.

5. Comparison of ΔG_{HX} and ΔG_U

The measured hydrogen-exchange stabilities of the most stable residues are generally higher than the conformational stabilities measured by urea and thermal denaturation, as shown for a variety of proteins in **Table 2**. There are two main reasons for the discrepancies: solvent isotope effects and proline isomerization effects.

5.1. Solvent Isotope Effect

To compare denaturation and hydrogen-exchange experiments, the traditional techniques must be performed in D_2O. At this time, we cannot predict whether a protein will be more or less stable in D_2O than in H_2O, since many factors affect the solvent isotope effect of a protein ***(30)***. We have ample evidence from data by Makhatadze et al. ***(31)***, the review by Oas and Toone ***(30)***, and Huyghues-Despointes et al. ***(12)*** that ΔG_U values can differ by ± 0–2 kcal/mol in H_2O compared to D_2O, and the difference may be dependent on pH.

5.2. Proline Isomerization Effect

Denatured states at native-state conditions in the hydrogen-exchange experiments differ from denatured states under equilibrium unfolding conditions in traditional experiments. Most peptide bonds in a folded protein are in a *trans* conformation. Only 0.03%

Table 2
Comparison of Conformational Stabilities Measured by Hydrogen Exchange and Traditional Methods[a]

Protein	No. of Proline Residues		ΔG_{HX} (kcal/mol)	ΔG_{HX}* (kcal/mol)	$\Delta G_U(D_2O)$ (kcal/mol)
	cis	*trans*			
RNase T1	2	2	10.7	8.2	7.9 (UDC) 8.0 (TDC)
A21G[b]				7.2	7.1 (UDC)
G23A[b]				7.1	7.0 (UDC)
A21G+G23A[b]				6.3	6.0 (UDC)
RNase A	2	2			
Mayo and Baldwin			(10.2)	7.8	7.7 (DSC)
Wang et al.			10.3	8.0	8.1 (DSC)
Neira et al.			9.4	7.0	7.0 (DSC)
barnase	0	3	8.3	8.1	8.3 (DSC)
CI2	0	5	7.6	7.1	7.0 (GDC)
434 cro	0	2	4.0	3.9	3.7 (UDC)
apocytochrome b_{562}	0	4	5.5	5.3	5.0 (DSC) 3.3 (GDC)
cytochrome c (equine)	0	4			
Bai et al.			(9.3)	9.0	9.1 (DSC)
Foord and Leatherbarrow			6.4	6.1	6.1 (DSC)
cytochrome c (yeast)	0	4	6.8	6.5	6.4 (TDC)
HEWL	0	2	12.4	12.2	11.7 (DSC)
NTL9	0	1	4.8	4.7	4.7 (TDC+UDC)
OMTKY3	1	2	8.2	7.2	7.2 (DSC)
PPL	0	0	(6.9)	6.9	6.7 (TDC) 4.9 (GDC)
HPr (*Escherichia coli*)	0	2	5.8	5.7	4.7 (UDC)
src SH3 domain	0	2	6.2	6.1	4.7 (TDC) 4.7 (GDC)
Barstar	0	2	6.2	6.0	5.0 (GDC)
RNase H* (*Escherichia coli*)	1	4	(10.9)	9.3	9.9 (GDC)
RNase H (*T. thermophilus*)	1	11	(15.8)	13.8	13.3 (GDC)
CTXIII[c]		5	6.6	6.3	6.0 (GDC)
RNase A[d]	2	2	11.8	9.6	9.2 (DSC)
RNase Sa[e]	1	5	8.9	7.3	7.5 (GDC)
T4 lysozyme[f]		3	(17.7)	17.5	16.0 (GDC)
SNase[g]	2	4			
Unligated			6.4	4.4	6.1 (UDC)
Ligated (Ca^{2+} and pdTp)			8.2	6.2	7.4 (UDC)
SNase[h] (Unligated)	2	4	~ 6.4	~ 4.5	4.0 (GDC)
E75A			~ 4.4	~ 2.5	2.0 (GDC)

(Continued)

Table 2 ***(continued)***

Protein	No. of Proline *cis*	Residues *trans*	ΔG_{HX} (kcal/mol)	ΔG_{HX}* (kcal/mol)	$\Delta G_U(D_2O)$ (kcal/mol)
M26G			~ 4.4	~ 2.5	2.3 (GDC)
D77A			~ 4.1	~ 2.2	3.1 (GDC)
V23A			~ 3.4	~ 1.5	2.2 (GDC)

[a] Unless otherwise indicated below, the specific conditions for each protein are given in the footnote to **Table 1** in Huyghues-Despointes et al. ***(12)***. ΔG_{HX} is the average ΔG_{HX} value of the three most stable residues. ΔG_{HX} values in parentheses were estimated by extrapolation of hydrogen exchange stabilities at higher denaturant concentrations ***(15)***. $\Delta G_{HX}^* = \Delta G_{HX} - \Sigma\Delta G_{pro}$. $\Delta G_U(D_2O)$ is measured by the LEM (**Eq. 1**) using urea (UDC) or GdmCl (GDC) denaturation curves, or by the Gibbs-Helmholtz equation (**Eq. 2**) using thermal denaturation curves (TDC) or differential scanning calorimetry (DSC).

[b] pD = 7.4 and 25°C ***(27)***.

[c] Cardiotoxin analog III. pD = 3.6 and 25°C ***(39)***.

[d] pD = 6.4 and 25°C ***(40)***.

[e] pD = 5.9 and 30°C (D. Laurents, J.M. Pérez-Cañadillas, J. Santoro, M. Rico, D. Schell, C. N. Pace, and M. Bruix, personal communication).

[f] pD = 6.0 and 25°C ***(41)***.

[g] pD = 5.9 and 37°C (***42*** and references therein). Lys-Pro117 is 91% *cis* and His-Pro47 is 9% *cis* in the folded unligated protein. Lys-Pro117 is greater than 98% *cis* in the folded ligated protein.

[h] pD = 5.6 and 25°C ***(43)***.

of the nonproline bonds and 5.2% of the proline bonds are in a *cis* conformation in folded proteins ***(32)***. Conversely, an unfolded protein at equilibrium has nonproline bonds exclusively in the *trans* conformation and 6–38% of the proline bonds in the *cis* conformation, depending on the identity of the residue preceding proline. The latter values are based on data by Reimer et al. ***(33)*** on capped pentapeptides with 20 different amino acids at the Xaa-Pro position (*see* **Table 3**).

Transiently unfolded states under native conditions in the hydrogen-exchange experiments have the proline bonds in the same conformation as in the folded protein because the rate of refolding and hydrogen exchange is often much faster than the rate of proline

Table 3
Proline Isomerization Based on Ac-Ala-Xaa-Pro-Ala-Lys-NH_2[a]

Xaa	% cis[b]	K_{Pro}^{U}[c]	RT ln $(1+K_{Pro}^{U})$[d]	RT ln $(1 + 1/K_{Pro}^{U})$[e]
Pro	6.0	15.67	1.67	0.037
Lys	6.8	13.71	1.59	0.042
Arg	7.2	12.89	1.56	0.044
Asp	7.3	12.70	1.55	0.045
Ala	7.7	12.00	1.52	0.047
Cys	8.7	10.49	1.45	0.054
Glu	9.0	10.11	1.43	0.056
Thr	9.4	9.64	1.40	0.058
His(0)[f]	9.5	9.53	1.39	0.059
Met	10.0	9.00	1.36	0.062
Ser	10.3	8.71	1.35	0.064
Val	10.4	8.62	1.34	0.065
Gln	11.5	7.70	1.28	0.072
Asn	11.6	7.62	1.28	0.073
Leu	12.0	7.33	1.26	0.076
Ile	12.0	7.33	1.26	0.076
Gly	13.7	6.30	1.18	0.087
His(+)[g]	16.5	5.06	1.07	0.107
Phe	23.0	3.35	0.87	0.155
Tyr	24.0	3.20	0.85	0.163
Trp	37.7	1.65	0.58	0.280

[a] Replicated from **Table 2** in Huyghues-Despointes et al. ***(12)***.

[b] The percentage of *cis* conformer measured at 23°C in 20 m*M* sodium phosphate at pH 6.0 ***(33)***. This value is relatively independent of temperature ***(38)***.

[c] K_{Pro} is the equilibrium constant for cis ⇔ *trans* isomerization of the indicated Xaa-Pro residue.

[d] Correction in kcal/mol at 25°C (**Eq. 10**) for *cis* Xaa-Pro residues.

[e] Correction in kcal/mol at 25°C (**Eq. 11**) for *trans* Xaa-Pro residues.

[f] Measured at pH 8.0.

[g] Measured at pH 3.5.

isomerization ***(12)***. The protein typically spends less than 1 s in the transiently populated unfolded state before all the amide protons are exchanged. This is shown for the folding equilibrium of a protein with a proline residue in the sequence:

$$F_{cis} \xleftrightarrow{K_{HX}} U_{cis} \xleftrightarrow{K^{U}_{pro}} U_{trans} \tag{8}$$

where F_{cis} and U_{cis} are the native and unfolded states with a Pro residue exclusively in the *cis* conformation. K^{U}_{pro} is the equilibrium constant for *cis/trans* proline isomerization in the unfolded state, and K_{HX} is the equilibrium constant for folding exclusively in the *cis* conformation. The denatured states will have a higher free energy in the hydrogen-exchange experiments than under the traditional unfolding experiments where the proline bonds reach *cis/trans* equilibrium. ΔG_{HX} measures ΔG for the first step in the equilibrium, but ΔG_U measures ΔG for both steps ***(15,34)***. The total difference is the stability from proline isomerization in the folded and unfolded states:

$$\Delta G_{pro} = \Delta G_{HX} - \Delta G_U = RT \ln (1 + K^{U}_{pro})/(1 + K^{F}_{pro}) \tag{9}$$

The full expression (**Eq. 9**) must be used for proteins such as Staphylococcal nuclease (SNase) where two proline residues flip between *cis* and *trans* conformers in the folded state (*see* **Table 2**).

For most proteins, where the proline conformation is fixed in the native conformation, the stability from proline isomerization caused by a *cis* proline in the native state can be reduced to:

$$\Delta G^{U}_{pro} = RT \ln(1 + K^{U}_{pro}) \tag{10}$$

or using **Eq. 8**, a *trans* proline in the native state can be:

$$\Delta G^{U}_{pro} = RT \ln(1 + 1/K^{U}_{pro}) \tag{11}$$

The ΔG^{U}_{pro} values for a *cis* proline in the native state are more than 1 kcal/mol, while ΔG^{U}_{pro} for a *trans* proline are much smaller (< 0.3 kcal/mol). **Table 3** shows the ΔG^{U}_{pro} values for *cis* and *trans* proline conformers at 25°C. The total effect of proline isomerization is determined by summing ΔG^{U}_{pro} contributions for all *cis* and *trans* proline residues in a protein. The conformation of a Xaa-Pro bond is detected by measuring the ω angle in the best NMR or X-ray crystal structure of a protein.

Until recently, it was commonly assumed that the proline bond was approx 20% *cis* in the unfolded state at equilibrium for all Xaa-Pro residues in a protein. As a result, the stabilities of the most stable residues were not always considered to exchange by global unfolding (*see* **refs. *35–37***), because these stabilities did not agree and were often higher than the values from traditional methods.

6. Conclusion

If the corrections for the proline isomerization of the Xaa-Pro bond using values of Reimer et al. ***(33)*** are applied and if the traditional experiments are performed in D_2O, we find an excellent agreement between the corrected ΔG_{HX} and ΔG_U for most proteins (**Table 1**). This comparison leads to several important conclusions:

1. The most stable amide hydrogens generally exchange by global unfolding.
2. The unfolded states that exist under native conditions are thermodynamically equivalent to those that exist after solvent or thermal unfolding.
3. Chemical exchange rates from short peptides can model well k_{rc}, the exchange-rate constant of the open conformation.
4. The equilibrium constants for proline isomerization in capped pentapeptides can be used to correct for the higher free energy of the unfolded states in the hydrogen-exchange experiments.

7. A Note on Nomenclature

In the literature, the nomenclature for describing the hydrogen-exchange process varies widely. Examples include defining the chemical exchange-rate constants as k_{rc}, k_{ch}, or k_{int}, the hydrogen-exchange stabilities as ΔG_{HX}, ΔG_{HD}, or ΔG_{ex}, the conformational stability as ΔG_U, ΔG_W, or $\Delta G(H_2O)$, the hydrogen-exchange process as HX, HD, or hydrogen exchange, and the pH meter reading of the experiment in D_2O as pH*, pH_{read}, or pD. In this chapter, all these terms are defined as the following: the random coil rate as k_{rc}, the hydrogen exchange stabilities as ΔG_{HX}, the conformational stability in 0 *M* denaturant as $\Delta G_U(D_2O)$ in deuterated solvent and $\Delta G_U(H_2O)$ in protonated solvent, the exchange process as hydrogen-exchange or HX, and the pH meter reading as pH*. In the future, for the purpose of clarity, it would be advisable to universally adopt the same nomenclature, and we suggest that the nomenclature defined here would serve this purpose.

Acknowledgment

Work in the authors' laboratories is supported by the National Institutes of Health (J. M. Scholtz, C. N. Pace, and S. W. Englander) and the Robert E. Welch Foundation (J. M. Scholtz and C. N. Pace).

References

1. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J.* **76,** 75–83.
2. Pace, C. N. (1975) The stability of globular proteins. *CRC Crit. Rev. Biochem.* **3,** 1–43.
3. Freire, E. (1995) Thermal denaturation methods in the study of protein folding. *Methods Enzymol.* **259,** 144–168.
4. Pace, C. N. and Scholtz, J. M. (1997) Measuring the conformational stability of a protein, in *Protein Structure. A Practical Approach,* (Creighton, T. E., ed.), Oxford University Press Inc., New York, NY, pp. 299–321.
5. Jaenicke, R. (1987) Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49,** 117–237.
6. Makhatadze, G. I. (1999) Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *J. Phys. Chem.* **103,** 4781–4785.
7. Chamberlain, A. K. and Marqusee, S. (1997) Touring the landscapes: partially folded proteins examined by hydrogen exchange. *Structure* **5,** 859–863.
8. Hvidt, A. and Neilson, S. O. (1966) Hydrogen exchange in proteins. *Adv. Protein Chem.* **21,** 287–386.
9. Bai, Y., Milne, J. S., Mayre, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* **17,** 75–86.

10. Loh, S. N., Rohl, C. A., Kiefhaber, T., and Baldwin, R. L. (1996) A general two-process model describes the hydrogen exchange behavior of RNase A in unfolding conditions. *Proc. Natl. Acad. Sci. USA* **93,** 1982–1987.
11. Qian, H. and Chan, S. I. (1999) Hydrogen exchange kinetics of proteins in denaturants: a generalized two-process model. *J. Mol. Biol.* **286,** 607–616.
12. Huyghes-Despointe, B. M. P., Scholtz, J. M., and Pace, C. N. (1999) Protein conformational stabilities can be determined from hydrogen-exchange rates. *Nat. Struct. Biol.* **6,** 910–912.
13. Connelly, G. P., Bai, Y., Jeng, M.-F.. amd Englander, S. W. (1993) Isotope effect in peptide group hydrogen exchange. *Proteins* **17,** 87–92.
14. Molday, R. S., Englander, S. W., and Kallen, R. G. (1972) Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11,** 150–158.
15. Bai, Y., Milne, J. S. Mayne, L., and Englander, S. W. (1994) Protein stability parameters measured by hydrogen exchange. *Proteins* **20,** 4–14.
16. Roder, H., Wagner, G., and Wüthrich, K. (1985) Amide proton exchange in proteins by EX1 kinetics: studies of the basic pancreatic trypsin inhibitor at variable pD and temperature. *Biochemistry* **24,** 7396–7407.
17. Tüchsen, E. and Woodward, C. (1987) *Biochemistry* **26,** 8073–8078.
18. Swint-Kruse, L. and Robertson, A. D. (1996) Temperature and pH dependences of hydrogen exchange and global stability for ovomucoid third domain. *Biochemistry* **35,** 171–180.
19. Mayo, S. L. and Baldwin, R. L. (1993) Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* **262,** 873–876.
20. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native state hydrogen exchange. *Science* **269,** 192–197.
21. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nat. Struct. Biol.* **3,** 782–787.
22. Grantcharova, V. P. and Baker, D. (1997) Folding dynamics of the src SH3 domain. *Biochemistry* **36,** 15,685–15,692.
23. Yi, Q., Scalley, M. L., Simons, K. T., Gladwin, S. T., and Baker, D. (1997) Characterization of the free energy spectrum of peptostreptococcal protein L. *Fold. Des.* **2,** 271–279.

24. Bhuyan, A. K. and Udgaonkar, J. B. (1998) Two structural subdomains of barstar detected by rapid mixing NMR measurement of amide hydrogen exchange. *Proteins* **30,** 295–308.
25. Fuentes, E. J. and Wand, A. J. (1998) Dynamics and stability of apocytochrome b562 examined by hydrogen exchange. *Biochemistry* **37,** 3687–3698.
26. Fuentes, E. J. and Wand, A. J. (1998) Local stability and dynamics of apocytochrome B562 examined by the dependence of hydrogen exchange on hydrostatic Pressure. *Biochemistry* **37,** 9877–9883.
27. Huyghes-Despointe, B. M. P., Langhoest, U., Steyaert, J., Pace, C. N., and Scholtz, J. M. (1999) Hydrogen-exchange stabilities of RNase T1 and variants with buried and solvent-exposed Ala→Gly mutations in the helix. *Biochemistry*
28. Pace, C. N., Laurents, D. V., and Thomson, J. A. (1990) pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry* **29,** 2564–2572.
29. Glasoe, P. F. and Long, F. A. (1960) Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **64,** 188–193.
30. Oas, T. G. and Toone, E. J. (1997) Thermodynamic solvent isotope effects and molecular hydrophobicity, in *Adv. Biophys. Chem.* (Bush, C., ed.), JAI Press, Inc., Greenwich, CT pp. 1–52.
31. Makhatadze, G. I., Clore, G. M., and Gronenborn, A. M. (1995) Solvent isotope effect of protein stability. *Nat. Struct. Biol.* **2,** 852–855.
32. Jabs, A., Weiss, M. S., and Hilgenfeld, R. (1999) Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286,** 291–304.
33. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects of peptidyl-prolyl *cis/trans* isomerization. *J. Mol. Biol.* **279,** 449–460.
34. Sharp, K. A. and Englander, S. W. (1994) How much is a stabilizing bond worth? *Trends Biochem. Sci.* **19,** 526–529.
35. Kragelund, B. B., Knudsen, J., and Poulsen, F. M. (1995) Local perturbations by ligand binding of hydrogen deuterium exchange kinetics in a four-helix bundle protein, acyl coenzyme A binding protein (ACBP). *J. Mol. Biol.* **250,** 695–706.
36. Li, R. and Woodward, C. (1999) Hydrogen exchange and protein folding. *Protein Sci.* **8,** 1571–1591.
37. Neira, J. L., Sevilla, P., Menéndez, M., Bruix, M., and Rico, M. (1999) Hydrogen exchange in ribonuclease A and ribonuclease S:

evidence for residual structure in the unfolded state under native conditions. *J. Mol. Biol.* **285,** 627–643.

38. Sivaraman, T., Kumar, T. K. S., and Yu, C. (1999) Investigation of the structural stability of cardiotoxin analogue III from the taiwan cobra by hydrogen-deuterium exchange kinetics. *Biochemistry* **38,** 9899–9905.
39. Chakshusmathi, G., Ratnaparkhi, G. S., Madhu, P. K., and Varadarajan, R. (1999) Native-state hydrogen-exchange studies of a fragment complex can provide structural information about the isolated fragments. *Proc. Natl. Acad. Sci. USA* **96,** 7899–7904.
40. Llinás, M., et al. (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat. Struct. Biol.* **6,** 1072–1076.
41. Loh, S. N., Prehoda, K. E., Wang, J., and Markley, J. L. (1993) Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry* **32,** 11,022–11,028.
42. Wrabl, J. and Shortle, D. (1999) A model of the changes in denatured state structure underlying m value effects in staphylococcal nuclease. *Nat. Struct. Biol.* **6,** 876–883.
43. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerization of proline residues. *Biochemistry* **14,** 4953–4963.

5

Modeling the Native State Ensemble

Vincent J. Hilser

1. Introduction

While it has been known for years that within a protein's folding/unfolding transition region the conformational equilibrium can be well approximated by a "two-state" model ***(1)***, it is not possible within the context of this "two-state" framework to account for an important and well-established experimental observation. As shown by a number of researchers, the hydrogen-exchange protection factor (PF) patterns of all proteins observed to date show large regional differences ***(2–12)***. If protein folding is "two-state", all amides that are protected from exchange in the native state show identical protection factors, as they become exposed to solvent through the same folding/unfolding reaction. The heterogeneity observed in the protection-factor patterns suggests that the native state is best characterized as a fluctuating ensemble of conformations that involve local unfolding reactions. In an effort to derive a quantitative description of these fluctuations, we developed the COREX algorithm ***(13–15)***, wherein the native state is represented as an ensemble rather than a discrete conformational state. The purpose of this chapter is to present a detailed description of the COREX algorithm, and to show how this ensemble-based description provides a uniform framework for reconciling the cooperative folding/unfolding behavior observed

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

in the transition region with the apparently noncooperative behavior seen under native conditions. This approach also represents a powerful tool for investigating cooperativity and the propagation of mutational effects.

2. Statistical Thermodynamic Formalism

The probability of any given conformation of a protein under equilibrium conditions can be described by the equation:

$$P_i \frac{\exp\left(^{-\Delta G_i}\!/_{RT}\right)}{Q} \tag{1}$$

where exp(–ΔG_i/RT) is the statistical weight or Boltzmann exponent of state *i* expressed in terms of the relative Gibbs free energy (ΔG_i). The partition function, Q, is the sum of the Boltzmann exponents of all states that are accessible to the system:

$$Q = \sum_{i=0}^{N} \exp\left(^{-\Delta G_i}\!/_{RT}\right) \tag{2}$$

in which ΔG_i can be written in terms of the standard thermodynamic equation:

$$\Delta G_i = \Delta H_i - T \cdot \Delta S_i \tag{3}$$

where ΔH_i and ΔS_i are the relative enthalpy and entropy of state *i* at temperature T, and R is the gas constant. **Eqs. 1–3** indicate that knowledge of the relative energies of each state provides direct access to the equilibrium distribution of states in the ensemble.

3. The COREX Algorithm

According to the COREX algorithm ***(13–15)***, a protein is modeled as an ensemble of partially folded conformations, each of which is derived from the high resolution X-ray or NMR structure. The individual conformations are generated by unfolding the various regions of the protein in exhaustive combinations as described in

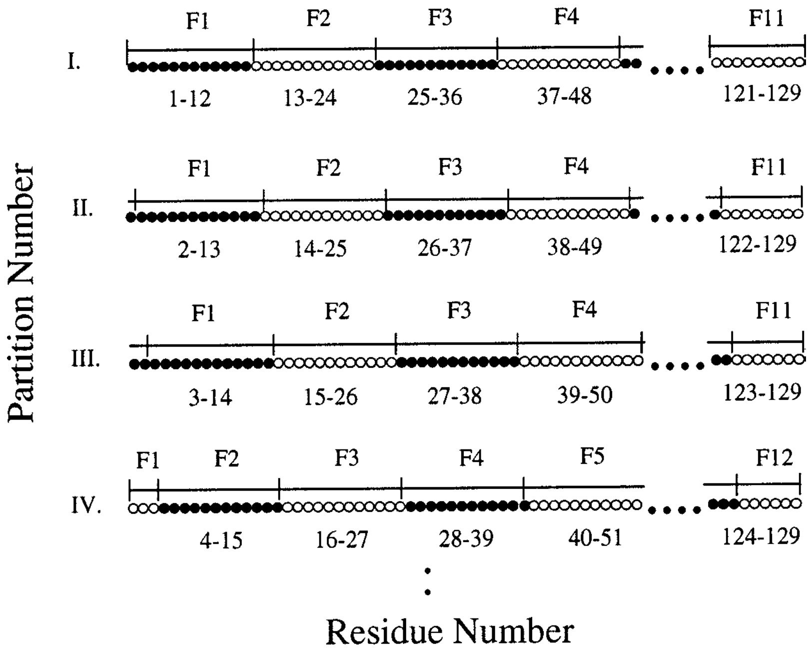


Fig. 1. An illustration of the method employed by the COREX algorithm for generating partially unfolded states. For this example, a protein of 129 amino acids is divided into 11 different sequential folding units (labeled F1–F11), each with 12 residues—except for the last unit, which has 9 residues, as shown in Partition I. The partitioning scheme is then overlaid on the high-resolution structure, and the different folding units are folded and unfolded in all possible combinations resulting in $2^N - 2$ (= 2046) partially folded states, where N is the number of folding units. To systematically vary the folding units, the boundaries are advanced one residue in the sequence as shown in Partition II. The new partition scheme is again overlaid on the high-resolution structure, and the combinatorial folding is repeated. In cases where less than three residues are present at the ends of the sequence, the residues are included in the adjoining folding unit. This procedure is repeated through 12 partitions, at which point the scheme repeats itself. The total number of states generated by this scheme is 32,757.

Fig. 1. The partition scheme shown in **Fig. 1** generates an ensemble of 32,757 different conformations.

From among these possible conformations, the most probable states are determined by calculating the free energy of each state using a structural energetic parameterization ***(16–24)***. In short, the parameterization allows us to calculate the heat capacity (ΔCp), enthalpy (ΔH), and entropy (ΔS) difference between each conformational state and the high-resolution structure (i.e., the reference state).

The heat-capacity change for each state has been parameterized in terms of changes in solvent-accessible polar ($\Delta ASApol$) and apolar ($\Delta ASAap$) surface area, as ΔCp is known to originate primarily from changes in hydration upon unfolding ***(16–18)***:

$$\Delta Cp = \Delta C_{p,ap} + \Delta C_{p,pol}$$

$$\Delta Cp = ac \cdot \Delta \mathrm{ASA}_{ap} + \mathrm{b}_c \cdot \Delta \mathrm{ASA}_{pol} + (c_c - b_c) \cdot \Delta \mathrm{ASA}_{OH} \qquad (4)$$

where $a_c = 0.45\ \mathrm{cal \cdot K^{-1} \cdot mol^{-1} \cdot Å^{-2}}$, and $b_c = 0.26\ \mathrm{cal \cdot K^{-1} \cdot mol^{-1} \cdot Å^{-2}}$, $c_c = 0.17\ \mathrm{cal \cdot K^{-1} \cdot mol^{-1} \cdot Å^{-2}}$. The third term accounts for the positive contribution of hydroxyl groups to the measured ΔCp ***(18)***. In reality, ΔCp has a temperature dependence that becomes significant above 80°C. For high-temperature calculations, additional terms are included in **Eq. 4** that account for this dependence ***(17)***.

The bulk of the enthalpy change of unfolding also scales with ΔASA changes ***(16,19)*** and at 60°C, the median unfolding temperature for the database of proteins, the enthalpy can be written as:

$$\Delta H_{gen}(60°C) = \mathrm{a}_H \cdot \Delta \mathrm{ASA}_{ap} + b_H \cdot \Delta \mathrm{ASA}_{pol} \qquad (5)$$

where $a_H = -8.44\ \mathrm{cal \cdot mol^{-1} \cdot Å^{-2}}$ and $b_H = 31.4\ \mathrm{cal \cdot mol^{-1} \cdot Å^{-2}}$.

For the entropy change of unfolding, the COREX algorithm includes two contributions, one resulting from changes in solvation, and the other caused by changes in conformational entropy ($\Delta S = \Delta S_{solv} + \Delta S_{conf}$). The solvation contribution can be written in terms of the apolar and polar contributions of ΔCp (**Eq. 4**) if the temperatures at which $\Delta S_{solv,ap} = 0$ and $\Delta S_{solv,pol} = 0$ ($T_{s,ap}$* and $T_{s,pol}$*) are known:

$$\Delta S_{solv} = \Delta S_{solv,ap} + \Delta S_{solv,pol}$$

$$\Delta S_{solv}(T) = \Delta C_{p,ap} \cdot \ln(T/T^*_{s,ap}) + \Delta C_{p,pol} \cdot \ln(T/T^*_{s,pol}) \qquad (6)$$

$T_{s,ap}$* has been shown to be ~385 K ***(23)*** while $T_{s,pol}$* is ~335 K ***(22)***.

Conformational entropies are calculated by considering three contributions:

$$\Delta S_{conf} = \Sigma\ \Delta S_{bu-ex} + \Sigma\ \Delta S_{ex-un} + \Sigma\ \Delta S_{bb} \tag{7}$$

1) $\Sigma\Delta S_{bu-ex}$ is the summed entropy change for all side chains that are buried in the fully folded state and become exposed in a partially unfolded state. 2) $\Sigma\ \Delta S_{ex-un}$ is the summed entropy change of solvent-exposed side chains upon unfolding of the peptide backbone. 3) $\Sigma\Delta S_{bb}$ is the backbone entropy change for residues that become unfolded in partially unfolded states. The magnitude of the entropy contributions for each amino acid have been determined from computational analysis of the probabilities of the different dihedral and torsion angles ***(21,22)***, and are reproduced in **Table 1**. Additional entropic contributions arise from disulfide bridges, which can be accounted for as described ***(25)***. The contributions listed here constitute >95% of the entropy of stabilization for most proteins. Contributions resulting from protonation, the binding of ligands, etc., can be estimated from knowledge of pK values, protonation enthalpies, and experimental pH. In estimating these contributions to the stability of each partially folded state, the difference between the experimental and calculated stability of the protein is determined under a given set of pH and ionic strength conditions. The corresponding entropy difference, which usually accounts for <3% of ΔS_{conf}, is then used to define an entropy-scaling factor. The conformational entropy of each partially folded state is then normalized to the scaling factor.

4. Residue-Specific Equilibria

The most important aspect of the ensemble-based approach is our ability to calculate the probability for each residue to be in a folded or an unfolded conformation ***(13–14)***. These residue-specific probabilities are used to define residue stability constants, $\kappa_{f,j}$, as

Table 1
Conformational Entropies for Amino Acids[a]

Amino acid	$\Delta S_{\text{bu–ex}}$	$\Delta S_{\text{ex–un}}$	ΔS_{bb}
Ala	0.00	0.00	4.1
Arg	7.11	–0.84	3.4
Asn	3.29	2.24	3.4
Asp	2.00	2.16	3.4
Cys	3.55	0.61	3.4
Gln	5.02	2.12	3.4
Glu	3.53	2.27	3.4
Gly	0.00	0.00	6.5
His	3.44	0.79	3.4
Ile	1.74	0.67	2.18
Leu	1.63	0.25	3.4
Lys	5.86	1.02	3.4
Met	4.55	0.58	3.4
Phe	1.40	2.89	3.4
Ser	3.68	0.55	3.4
Thr	3.31	0.48	3.4
Trp	2.74	1.15	3.4
Tyr	2.78	3.12	3.4
Val	0.12	1.29	2.18

[a] Entropy values are in cal·K^{-1}·mol^{-1}. $\Delta S_{\text{ex–un}}$ and $\Delta S_{\text{bu–ex}}$ are from ***(21–22)***. $\Delta S_{\text{bu–ex}}$ contribution for each residue is normalized by the fraction of exposed surface in the fully folded form.

$$\kappa_{f,j} = \frac{\Sigma P_{f,j}}{\Sigma P_{nf,j}} \tag{8}$$

where $\Sigma P_{f,j}$ and $\Sigma P_{nf,j}$ are the summed probabilities of all states in the ensemble in which residue *j* is either folded or unfolded, respectively. According to **Eq. 8**, residues with high-stability constants will be folded in the majority of highly probable states, while residues with low-stability constants will be unfolded in those states. This concept is depicted in **Fig. 2** for a hypothetical protein that has four significantly populated states (labeled I–IV). Four residues (*A–D*) are shown which are either folded *(filled circles)* or unfolded *(open circles)* in each conformation. For this example, residue *D*

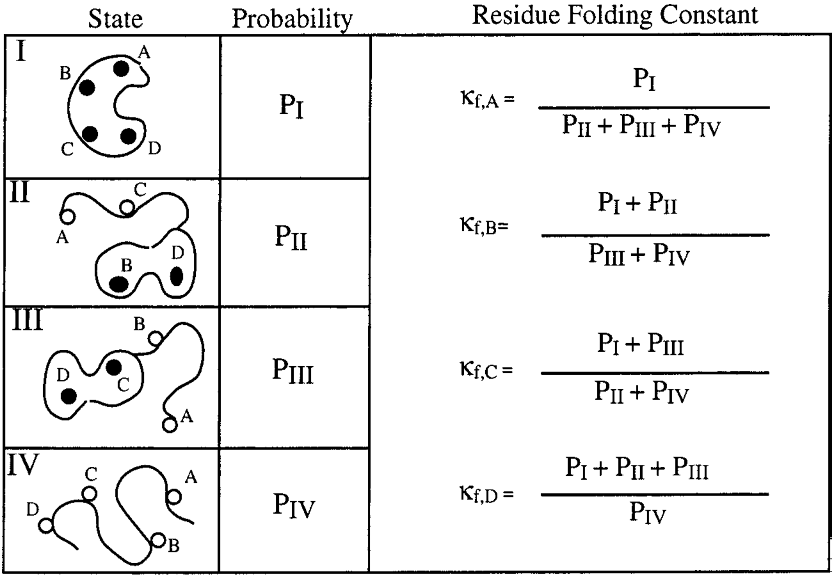


Fig. 2. Schematic representation of the residue stability constant for a hypothetical protein containing residues (*A*–*D*). For this example, four different conformational states (I–IV) are accessible wherein each residue is either folded (*closed circles*) or unfolded (*open circles*). The stability constant (**Eq. 8**) for each residue is the ratio of the summed probability of all states in which that residue is in a folded conformation to the summed probability that it is in an unfolded conformation.

will have the highest stability constant, as it is folded in the majority of the most probable states. Likewise, residue *A* will have the lowest stability constant.

5. Experimental Validation: Comparison with Hydrogen-Exchange Protection Factors

The significance of **Eq. 8** is that under certain experimental conditions, $\kappa_{f,j}$ for many residues, as calculated by the COREX algorithm, can be compared directly to protection factors (PF_j) determined from hydrogen/deuterium exchange (HX) experiments. Amide hydrogen exchange has become one of the most important tools in the analysis and quantitation of partially folded states of proteins ***(2–12)***. For those amide protons that are protected from exchange with solvent protons in the fully folded form of the protein, the exchange reaction can be described by:

$$\mathrm{NH_j(closed)} \xleftrightarrow{K_{op,j}} \mathrm{NH_j(open)} \xrightarrow{k_{int,j}} \mathrm{ND_j(exchanged)},$$

where open and closed refer to those states in which a particular amide is exchange-competent and exchange-incompetent, respectively. In the slow-exchange limit, whereby the intrinsic exchange rate of a particular amide, k_{int}, is much slower than the open-to-closed rate (i.e., in the bimolecular or EX2 exchange regime), the measured exchange can be expressed by:

$$k_{ex,j} = K_{op,j} \cdot k_{int,j} \tag{9}$$

Since k_{int} is known from the exchange rates of model peptides ***(26)***, the opening/closing equilibrium (K_{op}) can be determined directly from the measured exchange rate, k_{ex}. Thus, the degree of protection ($PF_j = 1/K_{op,j}$) provides a residue-specific free-energy difference between the exchange competent and incompetent species ($\Delta G_{j,closed/open} = -RT \cdot \ln PF_j$.)

Since PFs from HX measurements provide the opening/closing equilibrium and the stability constant provides a measure of the folding/unfolding equilibrium, comparisons of PF_j and $\kappa_{f,j}$ require recasting the PF in terms of the probability that a particular residue will be folded or unfolded:

$$\kappa_{f,j} = \frac{\Sigma P_{f,j} - \Sigma P_{f,xc,j}}{\Sigma P_{nf,j} + \Sigma P_{f,xc,j}} \tag{10}$$

where $\Sigma P_{f,xc,j}$ is the probability that a residue is folded, yet exchange-competent ***(14)***. As **Eq. 10** indicates, not all residues that are in a folded conformation are protected from exchange, and as a consequence, the expression for the PF differs from the residue stability constant (**Eq. 8**). Since the COREX algorithm allows us to calculate the probability of a residue to be folded yet exchange-competent, it is possible to calculate protection factors for all amides and compare them to those values that have been obtained experimentally.

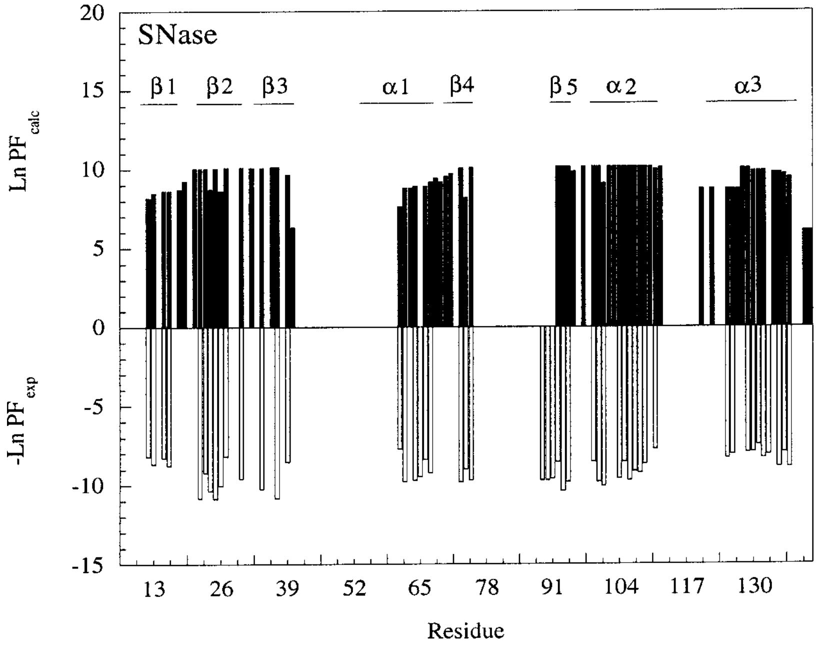
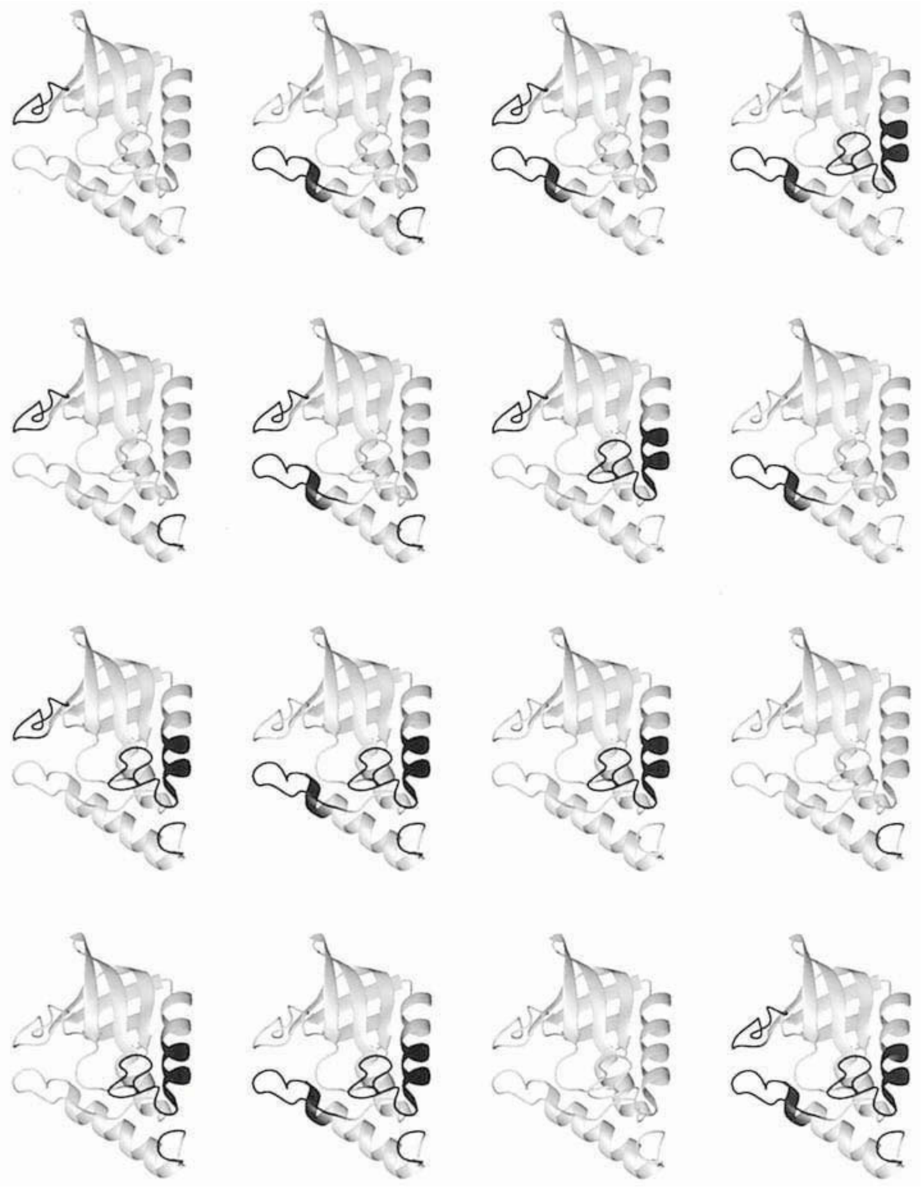


Fig. 3. Calculated and experimental protection factors for staphylococcal nuclease. Also shown are the locations of secondary structural elements. (Adapted from **ref.** ***14***).

Shown in **Fig. 3** is a comparison between the experimental and calculated PFs for staphylococcal nuclease (SNase) ***(4,13–14)***. As is the case for all proteins, SNase is characterized by regions of high- and low-stability constants. For SNase, regions with the highest stability constants lie in the β-barrel, while the loops between the elements of regular secondary structure have the lowest stability constants. In general—and not surprisingly—residues found in the core of the protein have the highest stability constants—they become unfolded only as a result of large-scale unfolding of the molecule. Residues in loop regions, on the other hand, become unfolded as a result of more local unfolding, which has a higher probability under native conditions. The good agreement between theory and experiment for SNase ***(13–14)***, λ_{6-85} ***(12)***, hen lysozyme ***(13)***, equine lysozyme ***(13)***, T4 lysozyme ***(15)***, BPTI ***(13)***, ovomu-



coid third domain ***(13)*** and ABL SH3 domain ***(27)*** suggests that the calculated ensemble generated by the COREX algorithm captures the general features of the actual ensemble.

Inspection of the most probable conformational states of the SNase ensemble as shown in **Fig. 4** reveals two significant features. First, the ensemble is biased toward states which have relatively small portions of the structure unfolded. These largely independent unfolding events are what give rise to the peaks and valleys in the PF patterns (**Fig. 3**). Second, although the high-resolution structure may provide the best average conformational state for any single region of the protein, the most probable states have some portion of the molecule disordered. The extremely low probability for the fully folded state of SNase is supported by rapid HX experiments using saturation-transfer techniques that show protection factors less than 1 for many residues ***(28)***. This result for SNase and other proteins highlights the potential pitfall of building a deterministic model from statistical data.

6. Denaturant Dependence of Stability Constants (PFs)

One of the hallmark features of native-state hydrogen-exchange studies has been a relatively diverse and complicated pattern of denaturant dependencies for different residues ***(8,29)***. In particular, it has been shown that at low denaturant concentrations, the PFs for many residues do not vary with denaturant, as shown for SNase in **Fig. 5**. As denaturant is increased, however, the PFs of all residues begin to converge to similar values. The origin of these effects can be elucidated by expanding **Eq. 8** to consider separately the fully folded (P_N) and fully unfolded (P_U) state contributions:

Fig. 4. *(opposite page)* Schematic representation of the ensemble of SNase conformations. White corresponds to regions that are folded, and black corresponds to regions that are unfolded in each conformational state. All states shown have different probabilities, but are within 1.0 kcal/mol of the minimum energy. Similar structures with slightly different borders are not shown. Fig. 4 was prepared with the program MOLMOL ***(34)***.

coid third domain ***(13)*** and ABL SH3 domain ***(27)*** suggests that the calculated ensemble generated by the COREX algorithm captures the general features of the actual ensemble.

Inspection of the most probable conformational states of the SNase ensemble as shown in **Fig. 4** reveals two significant features. First, the ensemble is biased toward states which have relatively small portions of the structure unfolded. These largely independent unfolding events are what give rise to the peaks and valleys in the PF patterns (**Fig. 3**). Second, although the high-resolution structure may provide the best average conformational state for any single region of the protein, the most probable states have some portion of the molecule disordered. The extremely low probability for the fully folded state of SNase is supported by rapid HX experiments using saturation-transfer techniques that show protection factors less than 1 for many residues ***(28)***. This result for SNase and other proteins highlights the potential pitfall of building a deterministic model from statistical data.

6. Denaturant Dependence of Stability Constants (PFs)

One of the hallmark features of native-state hydrogen-exchange studies has been a relatively diverse and complicated pattern of denaturant dependencies for different residues ***(8,29)***. In particular, it has been shown that at low denaturant concentrations, the PFs for many residues do not vary with denaturant, as shown for SNase in **Fig. 5**. As denaturant is increased, however, the PFs of all residues begin to converge to similar values. The origin of these effects can be elucidated by expanding **Eq. 8** to consider separately the fully folded (P_N) and fully unfolded (P_U) state contributions:

Fig. 4. *(opposite page)* Schematic representation of the ensemble of SNase conformations. White corresponds to regions that are folded, and black corresponds to regions that are unfolded in each conformational state. All states shown have different probabilities, but are within 1.0 kcal/mol of the minimum energy. Similar structures with slightly different borders are not shown. Fig. 4 was prepared with the program MOLMOL ***(34)***.

of the relative probabilities of states in the numerator and denominator of **Eq. 11**, and the capacity of each state to bind denaturant. In general, the ability of a state to bind denaturant and become stabilized depends on the amount of accessible surface area change (ΔASA); the greater the ΔASA, the more that state will be stabilized. Assuming a linear denaturant dependence of stability for each of the states in the ensemble ***(30)***,

$$\Delta G_i(Den) = \Delta G_i^o - m_i \cdot [Den] \tag{12}$$

where ΔG^o_i is the stability of state *i* in the absence of denaturant, and m_i is the characteristic slope for that state, it follows that:

$$m_i \propto \Delta \mathrm{ASA}_i$$

Thus, denaturant dependence of the stability of each state in the ensemble is related to the ΔASA for that state.

The denaturant dependence of the stability constant, on the other hand, does not provide the free-energy increment for a unique state. Taking the derivative of **Eq. 8** we obtain:

$$m_{app,j} = -RT \cdot \frac{\partial \ln \kappa_{f,j}}{\partial [Den]} = \langle m_{f,j} \rangle - \langle m_{nf,j} \rangle \tag{13}$$

where $<m_{f,j}>$ and $<m_{nf,j}>$ are the average denaturant dependence (i.e., $<m> = \Sigma P \cdot m$) for the subensembles in which residue *j* is folded and unfolded, respectively ***(13,14,31)***. **Equation 13** shows that the observed residue-specific denaturant dependence is the difference between the dependence of two subensembles. If both subensembles have the same <m> and thus <ΔASA> (=ΣP·ΔASA), then $<m_{f,j}> = <m_{nf,j}>$ and $m_{app} = 0$. This does not mean that those residues exchange as a result of unfolding that exposes no surface area. To the contrary, **Eq. 13** shows that any value of <ΔASA> can give zero for m_{app}, provided the subensembles have the same <ΔASA>. This point is highlighted in **Fig. 5**, which shows the calculated denaturant dependence of the stability constants for selected residues in SNase. As the data indicates, many of the residues show a flat denaturant dependence (i.e., $m_{app,j} = 0$) even though the stability constants for most of the residues shown in **Fig. 5** were derived from states which have a minimum of 9% (12/135) of the structure unfolded. This result emphasizes the statistical nature of the equilibrium, and demonstrates that relationships which hold true for macroscopic behavior (i.e., $m_{app} \alpha \Delta ASA$) do not necessarily apply to microscopic quantities.

7. Properties of the Native-State Ensemble

The COREX analysis has been applied to a large and diverse set of proteins ***(12–15,27)***. Where HX data is available, there is good agreement between the calculated and experimental protection factor patterns, suggesting that the calculated ensembles provide an accurate representation of the actual ensemble. Through inspection of the calculated ensemble, it becomes possible to elucidate the determinants of partially folded states. As noted, the ensembles of all proteins are characterized by states in which small segments of the structure are unfolded. For discussion purposes, the free energy of the partially folded states can be divided into three main categories: 1) a conformational entropy increment associated with the unfolding of a specific region; 2) a solvent-related effect originating from the unfolded region; and 3) a solvent-related effect originating from folded regions of the protein that become exposed as a result of unfolding elsewhere. The stability of each state is thus determined by the balance of these three contributions. In general, states with the highest probability have the least contributions from the third component ***(33)***. In other words, the unfolding of regions that do not result in a significant exposure of the hydrophobic core will have the highest probability. These regions usually correspond to loops that upon unfolding do not expose considerable amounts of surface.

8. Cooperativity

Since the calculated ensemble provides an accurate representation of the actual ensemble, it can be used as a tool to investigate a number of biological phenomena. Here we show that the ensemble-based formalism can be used to elucidate the origins and manifestations of cooperative behavior in proteins. The success of the COREX algorithm at reproducing the actual ensemble suggests that the approach correctly models—albeit implicitly—the network of cooperative interactions in proteins. This network of interactions determines the susceptibility of each region of the protein to changes at all other regions ***(12)***.

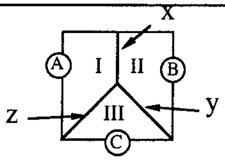
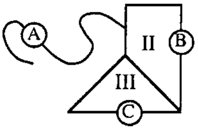
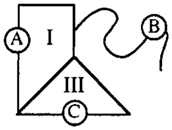
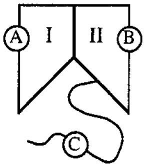
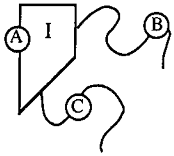
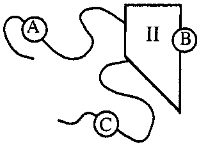
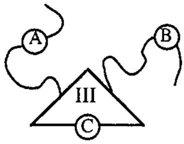
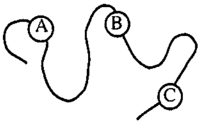
As the subset of states in which a particular residue is folded (or unfolded) differs for each residue (i.e., the summations in the numerator and denominator of **Eq. 8** are residue-specific), a perturbation to a specific residue, which changes the distribution of conformational states, will necessarily result in a different effect for each residue in the protein. This point is illustrated in **Fig. 6**, which shows a hypothetical protein containing three domains (labeled I, II and III) which interact through the surfaces *x*, *y*, and *z* as shown. Shown also in **Fig. 6** are the free energies and statistical weights of the potential states (i.e., all possible folding combinations for three domains). By introducing one residue to each domain (labeled *A*, *B*, and *C*, respectively), and deriving expressions for the stability constants:

$$\kappa_{f,A} = \frac{P_N + P_2 + P_3 + P_4}{P_U + P_1 + P_5 + P_6} \quad \kappa_{f,B} = \frac{P_N + P_1 + P_3 + P_5}{P_U + P_2 + P_4 + P_6} \quad \kappa_{f,C} = \frac{P_N + P_1 + P_2 + P_6}{P_U + P_3 + P_4 + P_5} \qquad (14A)$$

it is clear that the stability constants differ for each residue.

To demonstrate how cooperativity is manifested in an ensemble, we can consider a set of conditions within which the breaking of interaction *x* is energetically unfavorable while breaking interactions *y* and/or *z* are accompanied by no energetic cost (e.g, $\Delta G_I = \Delta G_{II} = -3.0$ kcal/mol, $\Delta G_{III} = 4.0$ kcal/mol, $\Delta g_x = 10.0$ kcal/mol, and $\Delta g_y = \Delta g_z = 0.0$ kcal/mol). For this simplified case, states 1, 2, 4, 5, and U will be energetically unfavored, and hence, will have probabilities that are much lower than states N, 3 and 6. As a result equation 14A reduces to:

$$\kappa_{f,A} \approx \frac{P_N + P_3}{P_6} \qquad \kappa_{f,B} \approx \frac{P_N + P_3}{P_6} \qquad \kappa_{f,C} \approx \frac{P_N + P_6}{P_3} \qquad (14B)$$

No.	State	ΔG_i	Stat. Wt.
N		0	1
1		$\Delta G_I + \Delta g_x + \Delta g_z$	$K_I \phi_x \phi_z$
2		$\Delta G_{II} + \Delta g_x + \Delta g_y$	$K_{II} \phi_x \phi_y$
3		$\Delta G_{III} + \Delta g_y + \Delta g_z$	$K_{III} \phi_y \phi_z$
4		$\Delta G_{II} + \Delta G_{III}$ $+ \Delta g_x + \Delta g_y + \Delta g_z$	$K_{II} K_{III} \phi_x \phi_y \phi_z$
5		$\Delta G_I + \Delta G_{III}$ $+ \Delta g_x + \Delta g_y + \Delta g_z$	$K_I K_{III} \phi_x \phi_y \phi_z$
6		$\Delta G_I + \Delta G_{II}$ $+ \Delta g_x + \Delta g_y + \Delta g_z$	$K_I K_{II} \phi_x \phi_y \phi_z$
U		$\Delta G_I + \Delta G_{II} + \Delta G_{III}$ $+ \Delta g_x + \Delta g_y + \Delta g_z$	$K_I K_{II} K_{III}$ $\bullet\ \phi_x \phi_y \phi_z$

where it is evident that residues *A* and *B* behave cooperatively as the expressions for their stability constants are approximately identical (**Fig. 7**). Thus, within the context of an ensemble approach, cooperativity between two residues is observed when the most probable states have both residues either folded or unfolded.

The development presented here provides clues as to how cooperativity can be accessed experimentally. If one were to design the ideal experiment for probing cooperativity in proteins, one would choose a technique in which the stabilities of specific regions of the protein are affected, but which result in no structural perturbation. The approach that most closely approximates the ideal case is to compare GLY and ALA mutants at surface-exposed sites throughout a test protein. In such cases, no structural perturbations would occur in the native state, and the stability change would arise almost entirely from differences in degrees of freedom in those states in which that residue is unfolded ***(22)***.

Such a scenario can be modeled using the hypothetical protein in **Fig. 6**. According to the ensemble view, the effect of a point mutation is to redistribute the conformational ensemble in a case-specific way. An GLY to ALA mutation at residue *A* which decreases the stability of all states in which residue *A* is unfolded (i.e., states 1, 5, 6, and U) will affect the stability constants as follows:

$$\kappa_{f,A} = \frac{P_N + P_2 + P_3 + P_4}{\psi_A\,(P_U + P_1 + P_5 + P_6)}$$

$$\kappa_{f,B} = \frac{P_N + P_3 + \psi_A(P_1 + P_5)}{\psi_A(P_U + P_6) + P_2 + P_4}$$

$$\kappa_{f,C} = \frac{P_N + P_2 + \psi_A(P_1 + P_6)}{\psi_A(P_U + P_5) + P_3 + P_4} \qquad (14C)$$

Fig. 6. *(opposite page)* Schematic representation of a hypothetical protein consisting of three domains (labeled I–III) which interact through surfaces *x*, *y*, and *z* as shown, and which contain residues *A*, *B*, and *C*. The free energies (ΔG_i) and statistical weights of all possible states are also shown. Upon unfolding any domain, the interaction energy between adjacent domains is broken ***(33)***.

where ψ_A (= $\exp(-\Delta g_{Mut\ A}/RT)$) is the magnitude of the effect of the perturbation at residue *A* on the probability of each state in which *A* is unfolded. As **Eq. 14C** indicates, the stability constant of each residue will be affected, but the magnitude of the effect to each residue will be determined by the relative probabilities of each state. In fact, two residues, which behave cooperatively under one set of conditions, may not behave cooperatively under a different set of conditions. **Figure 7** shows the residue stability constants and free energies ($\Delta G_j = -RT \cdot \ln \kappa_{f,j}$) for the energetic parameters listed in **Eq. 14C**. In the case of a mutation which decreases the stability of all states in which residue *A* is unfolded by 1.0 kcal/mol (i.e., $\Delta G_{Mut\ A}$, = 1.0 kcal/mol; $\psi_A = 0.18$), the result is a 5.41-fold change to $\kappa_{f,A}$, by definition ($\Delta\Delta G_A = -RT \cdot \ln \kappa_{f,A}$ = 1.0 kcal/mol), a 5.30-fold change in $\kappa_{f,B}$ ($\Delta\Delta G_B = -RT \cdot \ln \kappa_{f,B}$ = 0.98 kcal/mol), and no change in $\kappa_{f,C}$. As **Eq. 14B** suggests, this is the expected result, as residues *A* and *B* are cooperative while neither behaves cooperatively with respect to residue *C*.

To explore cooperativity in real proteins, the COREX algorithm may be applied to GLY and ALA mutations at surface-exposed sites ***(15)***. This is shown in **Fig. 8** for T4 lysozyme (T4Lys). In this example, two mutants of T4Lys were generated, representing ALA and GLY substitutions at the surface-exposed site S44 ***(34)***. As the side chain makes no intramolecular contacts in the native state, the crystallographic structures of the GLY and ALA variants differ by less than 0.14 Å backbone RMS ***(34)***. Inspection of the calculated ln κ_f values for the S44A (pdb 1L68) and S44G (pdb 107L) structures (**Fig. 8A**) reveals that despite the lack of appreciable structural differences, the effects of the mutations propagate to the entirety of the N terminal domain, as indicated in **Fig. 8B**. This effect stems from the fact that in the majority of highly probable states, all residues in the N terminal domain are either folded or unfolded. Therefore, a mutation affects all N-terminal-domain residues to a similar extent.

The results for T4Lys point toward an interesting paradox. In the context of a deterministic view, wherein proteins are viewed as single conformations, mutational effects are "sensed" by other regions of the protein via a structural deformation that propagates from the

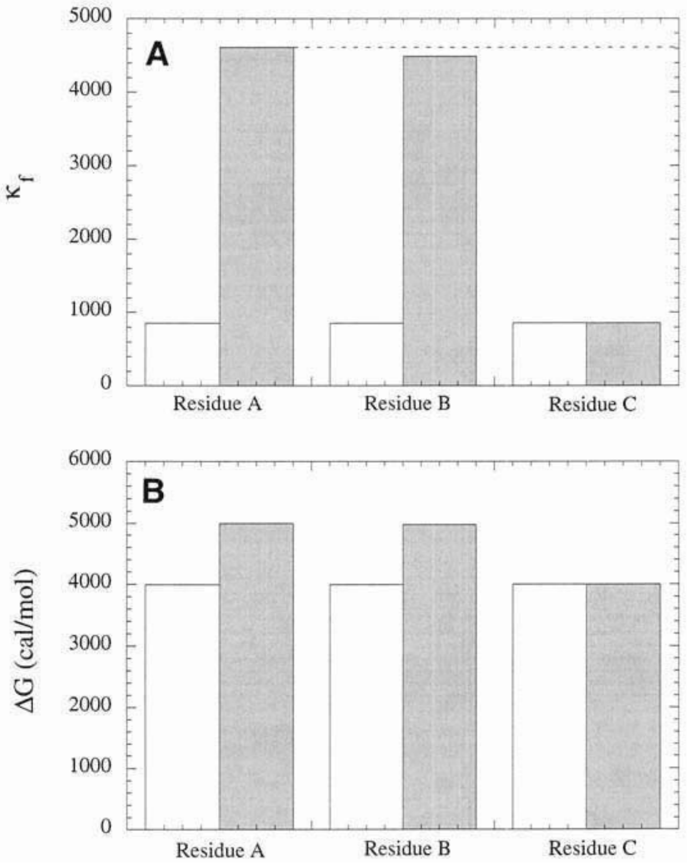


Fig. 7. Stability constants **(A)** and ΔG **(B)** for residues *A*, *B* and *C*. In the unperturbed protein (white bars) the energetic parameters are as follows: $\Delta G_I = \Delta G_{II} = -3.0$ kcal/mol, $\Delta G_{III} = 4.0$ kcal/mol, $\Delta g_x = 10.0$ kcal/mol, and $\Delta g_y = \Delta g_z = 0.0$ kcal/mol). A mutation of residue *A* (gray bars) which increases the stability of domain I by 1.0 kcal/mol ($\Delta G_I = -2.0$) will destabilize all states in which residue A is unfolded. The effect of such a change is residue-specific.

mutational site to the affected area. Such a mechanical view is an artifact imposed by the deterministic model. In the case of ALA/GLY mutations at residue 44 of T4Lys, the effect propagates as a result of a redistribution of the conformational ensemble. In solution-thermodynamic terms, the effect is almost entirely caused by a con-

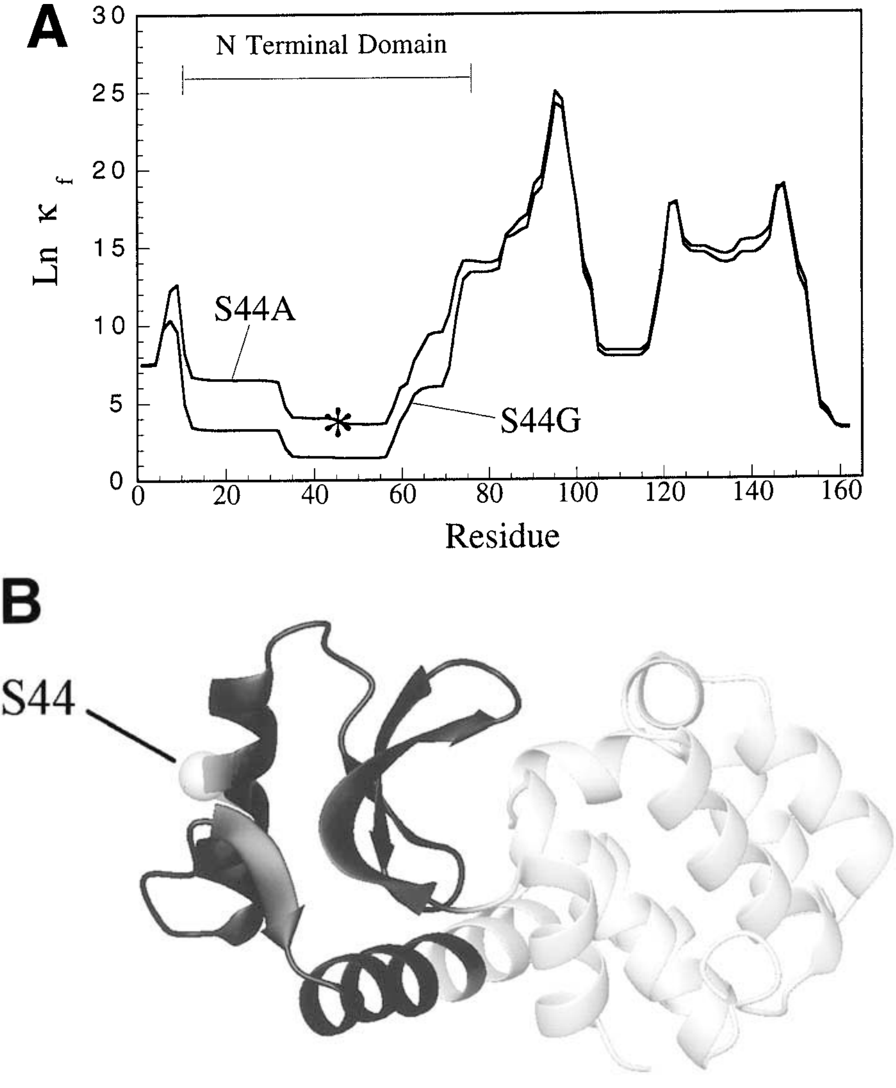


Fig. 8. **(A)** Residue stability constants (ln κ_f) for T4 lysozyme mutants S44A and S44G. Residue 44 is indicated by (*). Differences in stability constants are propagated to the entire N-terminal domain. **(B)** High-resolution structure of S44A T4 lysozyme (pdb 1L68). The N-terminal-domain (residues 13–71; colored black) is affected by the ALA to GLY change. The C-terminal-domain (residues 1–12 and 72–164; colored white) is largely unaffected. Fig. 8 was prepared with the program MOLMOL ***(36)***.

formational entropy change ***(15,22)***—an effect which is extremely difficult, if not impossible, to deduce from a deterministic model.

9. Conclusion

The formalism presented here inputs the high-resolution structure of the protein and generates, as an output, a large ensemble of states. As such, the analysis highlights the statistical nature of the equilibrium. Further, as this model is structure-based, it represents a link between experimental observables and statistical thermodynamics, and therefore can be used as a tool for interpreting results, deriving hypotheses, and directing research efforts. The significance of the ensemble-based formalism is twofold: First, it allows one to explain within a unified framework, the heterogeneity in protection factors under native conditions, the temperature and denaturant dependence of these values, and the cooperative folding/unfolding behavior seen experimentally ***(12–15)***. Second, it allows one to investigate cooperativity in proteins and, thereby show how perturbations in proteins are propagated through structure. The success of this model in describing cooperative behavior may have dramatic implications on our ability to explain allosterism and signal transduction ***(35)***, as these phenomena are merely more complex examples of how perturbations are propagated through structure.

References

1. Lumry, R., Biltonen, R., and Brandts, J. F. (1966) Validity of the “two-state” hypothesis for conformational transitions of proteins. *Biopolymers* **4,** 917–944.
2. Jeng, M.-F. and Englander, S. W. (1991) Stable submolecular folding units in a non-compact form of cytochrome c. *J. Mol. Biol.* **221,** 1045–1061.
3. Radford, S. E., Buck, M., Topping, K. D., Dobson, C. M., and Evans, P. A. (1992) Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins* **14,** 237–248.
4. Loh, S. N., Prehoda, K. E., Wang, J., and Markley, J. L. (1993) Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry* **32,** 11,022–11,028.

5. Kim, K.-S., Fuchs, J. A., and Woodward, C. K. (1993) Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry* **32,** 9600–9608.
6. Kim, K.-S. and Woodward, C. (1993) Protein internal flexibility and global stability: effect of urea on hydrogen exchange rates of bovine pancreatic trypsin inhibitor. *Biochemistry* **32,** 9609–9613.
7. Woodward, C. (1993) *TIBS* **18,** 359–360.
8. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* **269,** 192–197.
9. Schulman, B. A., Redfield, C., Peng, Z., Dobson, C. M., and Kim, P. S. (1995) Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human alpha-lactalbumin. *J. Mol. Biol.* **253,** 651–657.
10. Clarke, J. and Fersht, A. R. (1996) An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Folding and Design* **1,** 243–254.
11. Swint-Kruse, L. and Robertson, A. D. (1996) Temperature and pH dependences of hydrogen exchange and global stability for ovomucoid third domain. *Biochemistry* **35,** 171–180.
12. Hilser, V. J., Dowdy, D., Oas, T. G., and Freire, E. (1998) The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. USA* **95,** 9903–9908.
13. Hilser, V. J. and Freire, E. (1996) Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **262,** 756–772.
14. Hilser, V. J. and Freire, E. (1997) Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease. *Proteins* **27,** 171–183.
15. Hilser, V. J., Townsend, B. D., and Freire, E. (1997) Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: structural mapping of cooperative interactions. *Biophys. Chem.* **64,** 69–79.
16. Murphy, K. P. and Freire, E. (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* **43,** 313–361.
17. Gomez, J., Hilser, J. V., Xie, D., and Freire, E. (1995) The heat capacity of proteins. *Proteins Struct. Funct. Genet.* **22,** 404–412.

18. Habermann, S. M. and Murphy, K. P. (1996) Energetics of hydrogen bonding in proteins: a model compound study. *Protein Sci.* **5,** 1229–1239.
19. Xie, D. and Freire, E. (1994) Structure based prediction of protein folding intermediates. *J. Mol. Biol.* **242,** 62–80.
20. Hilser, V. J., Gomez, J., and Freire, E. (1996) The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins* **26,** 123–133.
21. Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994) Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins Struct. Funct. Genet.* **20,** 68–84.
22. DAquino, J. A., Gómez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25,** 143–156.
23. Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. USA* **83,** 8069–8072.
24. Luque, I., Mayorga, O., and Freire, E. (1996) Structure-based thermodynamic scale of alpha-helix propensities in amino acids. *Biochemistry* **35,** 13,681–13,688.
25. Pace, C. N., Grinsley, G. R., Thomson, J. A., and Barnett, B. J. (1988) Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem.* **263,** 11,820–11,825.
26. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* **17,** 75–86.
27. Sadqi, M., Casares, S., Abril, M. A., Lopez-Mayorga, O., Conejero-Lara, F., and Freire, E. (1999) The native state conformational ensemble of the SH3 domain from alpha-spectrin. *Biochemistry* **38,** 8899–8906.
28. Mori, S., van Zijl, P. C., and Shortle, D. (1997) Measurement of water-amide proton exchange rates in the denatured state of staphylococcal nuclease by a magnetization transfer technique. *Proteins Struct. Funct. Genet.* **28,** 325–332.
29. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Biol.* **3,** 782–787.
30. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131,** 266–280.

31. Wooll, J. O., Wrabl, J. O., and Hilser, V. J. (2000) Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *J. Mol. Biol.* **301,** 247–256.
32. Makhatadze, G. I. and Privalov, P. L. (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J. Mol. Biol.* **226,** 491–505.
33. Murphy, K. P., Bhakuni, V., Xie, D., and Freire, E. (1992) Molecular basis of co-operativity in protein folding. III. Structural identification of cooperative folding units and folding intermediates. *J. Mol. Biol.* **227,** 293–306.
34. Blaber, M., Zhang, X.-J., Lindstrom, J. L., Pepiot, S. D., Baase, W. A., and Matthews, B. W. (1994) Determination of alpha-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. *J. Mol. Biol.* **235,** 600–624.
35. Pan, H., Lee, J. C.. and Hilser, V. J. (2000) From the cover: binding sites in escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. USA* **97,** 12,020–12,025.
36. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14,** 51–55.

6

Conformational Entropy in Protein Folding

A Guide to Estimating Conformational Entropy via Modeling and Computation

Trevor P. Creamer

1. Introduction

This chapter outlines methods to estimate, through calculation and modeling, a major destabilizing contributor to protein stability—conformational entropy. This is not a comprehensive review of all published calculations of conformational entropy, but rather serves as a guide for those interested in performing such calculations. The methods outlined in this chapter provide an opportunity to estimate some of the entropic contributions to protein stability. Such methods are currently limited to estimating conformational entropy for relatively small parts of proteins, such as side chains or small peptides. Yet these calculations can provide much insight into the role of conformational entropy in protein stability.

1.1. Importance of ΔS_{conf} in Folding and Stability

Conformational entropy is one of the major contributors to the stability of proteins. It is one of the largest forces that opposes pro-

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

tein folding—and a force that may help to determine the structures of proteins. Conformational entropy is a measure of the degrees of conformational freedom available to a protein (or part thereof). The unfolded states of a protein possess an enormous number of degrees of freedom. A simple example to illustrate this is as follows: consider a relatively small, single domain protein of 100 residues in length. Let us assume that each residue in the protein can adopt two possible backbone conformations: one that corresponds to the α-helical region and one that corresponds to the β-region of (ϕ,ψ) space (a gross underestimation). Such a protein, when completely unfolded, could have up to 2^{100} available conformations (actually less because of excluded volume effects). This ignores side-chain conformations. Clearly, this relatively small protein possesses a vast degree of freedom, and thus a large amount of conformational entropy when unfolded. Almost all of its freedom is lost when the protein adopts its unique folded structure. Thus, protein folding results in an enormous loss of conformational entropy.

1.2. ΔS_{conf} in Secondary Structure Formation

It has been suggested that conformational entropy may be a guiding force in the formation of protein secondary structure ***(1)***. Several groups have demonstrated that loss of side-chain conformational entropy is an important contributor to the formation of α-helices ***(2–5)***. What contribution—if any—conformational entropy makes to the formation of β-strands and β-sheets is currently unknown. One suggestion is that β-strands arise when the side-chain entropy that would be lost upon helix formation is greater than the sum of the forces that stabilize helices ***(1)***. Even if conformational entropy does not guide the formation of β-sheet structures, it clearly still plays a role in the stability of such structures, since both side-chain and backbone entropy are lost upon folding. The methods for estimating conformational entropy described in this chapter are well-suited to exploring the role of entropy in secondary structure formation and stability.

2. Calculation of ΔS

2.1. Boltzmann's Equation

The entropy of a molecular system at equilibrium can be calculated using methods derived from statistical mechanics. If we know the partition function for a system, we can then calculate all thermodynamic properties of that system, including the entropy. As noted in many physical chemistry texts, the entropy of a molecular system, S, can be calculated using Boltzmann's formalism:

$$S = -R\sum_i p_i \ln p_i \tag{1}$$

where R is the gas constant, p_i is the probability of the system being in state i, and the sum is taken over all possible states of the system. Knowing the probabilities of being in each of the possible states of a system is equivalent to knowing the partition function for that system. If the states under consideration are conformational states, then the entropy calculated is the conformational entropy.

As illustrated by a simple example, calculation of the entropy is straightforward when the distribution of states is known. Suppose we have a system that consists of 10 atoms and that these atoms have three discrete energy states—A, B, and C (*see* **Fig. 1**). Let us assume in this example that the atoms are distributed so that, on average, two are in state A, three are in state B, and the remaining five are in state C (i.e., $\mathrm{E_A > E_B > E_C}$). The entropy of this system is then:

$$\begin{aligned} S &= R[\,p_A \ln p_A + p_B \ln p_B + p_C \ln p_C] \\ &= -R[0.2\ln 0.2 + 0.3\ln 0.3 + 0.5\ln 0.5] \qquad (2) \\ &= 1.03R \end{aligned}$$

2.2. Distributions

Clearly, to estimate the conformational entropy of a molecular system, we need the distribution of conformations for that system. In the case of protein folding where we are attempting to estimate

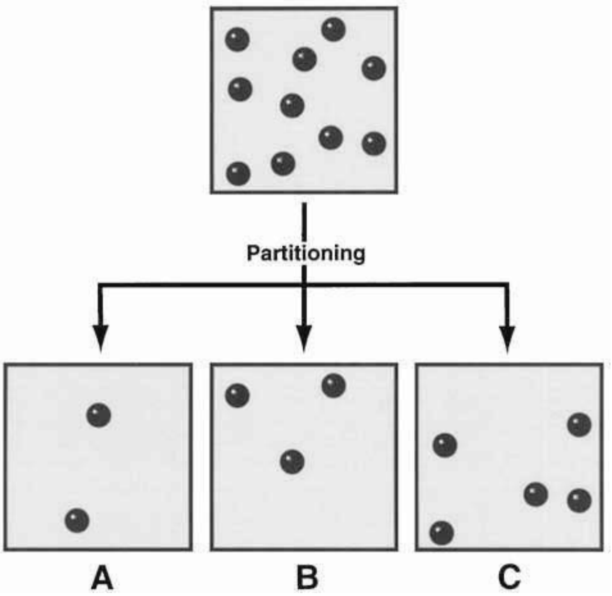


Fig. 1. Illustration of the use of Boltzmann's entropy equation.

ΔS_{conf} contributions to folding energetics, we need two distributions: the distribution of conformations in the folded state and the distribution in the unfolded state. The remainder of this chapter focuses on a discussion of how to generate or estimate these distributions.

2.3. Partitioning

Before discussing how to estimate the conformational distributions required for calculating ΔS_{conf}, it is important to discuss the partitioning of these data. The conformational distributions of proteins, or parts of proteins such as side chains, are potentially continuous. For example, rotation about a C–C bond within a side chain could lead to a distribution of χ torsion angles as shown in **Fig. 2.** In this case, all values of χ are possible and, given a sufficiently large data set, all angles could be populated. The question is: how do you partition such a distribution into discrete states in order to calculate the conformational entropy using **Eq. 1**?

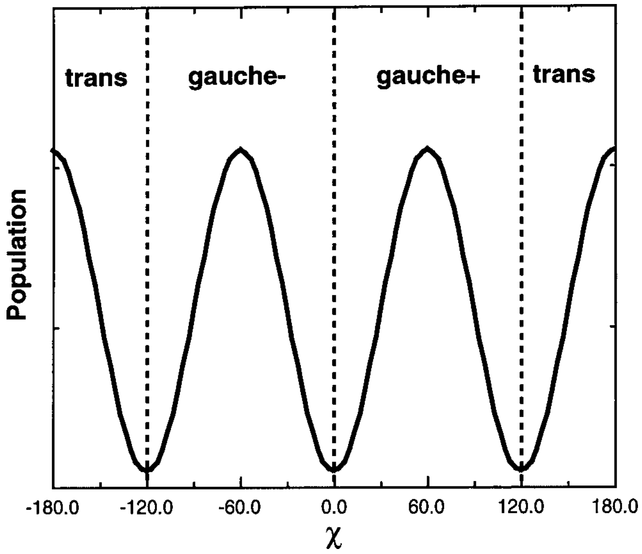


Fig. 2. Possible partitioning of a hypothetical conformational (rotamer) distribution for a single side-chain torsion.

One common way to partition conformational data is to assume that high-energy/low-population conformations contribute little, and that all similar conformations can be grouped together. In practice, these are reasonable assumptions, and this type of approach has been taken by a number of groups *(2–7)*. For the single side-chain torsion illustrated in **Fig. 2**, this could be done in the following manner: all conformations with $\chi < -120°$ or $\chi > +120°$ would be placed together in a group called *trans*, a second grouping called *g*- (gauche-) would consist of all conformations where $-120° < \chi < 0°$, and a third, *g*+ (gauche+) would be drawn from $0° < \chi < +120°$. Thus, the conformations are partitioned into three groupings, each spanning 120° (**Fig. 2**). The conformational entropy would then be estimated from:

$$S_{conf} = -R[p_{trans} \ln p_{trans} + p_{g-} \ln p_{g-} + p_{g+} \ln p_{g+}]. \quad (3)$$

In this example, the three groupings are centered around the three peaks in population. Although it would be easy to choose more than three groupings, this would lead to a different value for S_{conf}. In prior applications of this type of partitioning, conformational distributions have generally been split into groupings centered on peaks in the population of conformations ***(2–7)***. Distributions of populations can be obtained from a number of sources, including databases of known protein structures or Monte Carlo computer simulations.

A second approach to partitioning is to use the energy distribution rather than populations. The probability of a given conformation can be calculated using the energy of that conformation and the partition function for the system, Z. Thus, the probability of conformation *i* is given by:

$$p_i = \frac{e^{-E_i/RT}}{Z} \tag{4}$$

where E_i is the energy of conformation *i*, *R* is the gas constant, and *T* is the temperature. The partition function *Z* is calculated from:

$$Z = \sum_k e^{-E_k/RT} \tag{5}$$

where the sum is taken over all conformations *k* of the system. Energy distributions can be obtained from exhaustive conformational search methods. This approach has been used to estimate conformational entropy differences for pairs of side chains in α-helices ***(8)*** as well as backbone conformational entropy differences ***(9)***.

3. Conformational/Energetic Distributions

3.1. Introduction

There are a number of ways to obtain the conformational or energetic distributions required for estimating conformational entropy. The simplest approach is to obtain conformational distributions from a dataset of known protein structures. Other approaches involve generating the distributions *de novo* using computational methods. Each of these approaches is discussed in **Subheadings 3.2.–3.6.**

3.2. Conformational Distributions from Known Structures

As noted in **Subheading 3.1.**, this is the simplest approach. The first step is to assemble a representative dataset of high-resolution protein structures from the Protein Data Bank (PDB) ***(10)***. Hobohm and Sanders ***(11,12)*** provide a detailed discusion of how this is done. It is important to assemble the largest nonredundant set of high-resolution structures possible for the reasons outlined here. The second step is to extract from this dataset all occurrences of the residues/structures of interest, and to partition these into conformations. The conformational entropy can then be calculated using **Eq. 1**. This approach has been used to estimate scales of side-chain conformational entropy in the unfolded states of proteins ***(6,7,13)*** and to estimate the amount of conformational entropy lost by side chains when folded into α-helices ***(5)***.

The conformational entropy lost by a valine side chain upon folding into an α-helix is used as an example to illustrate this approach. Blaber et al. ***(5)*** assembled a dataset of 100 high-resolution protein structures. There were 158 valines in internal positions within α-helices and 781 in nonhelical conformations within this dataset. The conformational distributions for these valines are provided in **Table 1**. Assuming that the distribution of valines in nonhelical positions approximates the distribution in the unfolded states of proteins, one can then estimate the loss of side-chain conformational entropy for a valine folded into an internal position in an α-helix from the data in **Table 1**, leading to $-T\Delta S = 0.27$ kcal.mol.$^{-1}$ ***(5)***.

The major problem with this approach is the lack of data. In the example given here, there were 158 valines in internal α-helical positions. Since there are three major conformations for valine—*g*–, *trans* and *g*+—this is probably sufficient data to describe the conformational distribution. Lysine has four side-chain torsions (ignoring hydrogens), each of which can be partitioned into three positions, *g*–, *trans* and *g*+, leading to $3^4 = 81$ possible conformations. In the 100-protein dataset used by Blaber et al. ***(5)***, there are

Table 1
Conformational Distributions of Valine Side Chains from a Dataset of 100 High-Resolution Protein Structures

	Nonhelical residues		Internal helical residues	
Rotamer	Number	Percent	Number	Percent
	781		158	
Gauche–	202	26	10	6
Gauche+	68	9	6	4
Trans	511	65	142	90

Adapted from Blaber et al., ***(5)***.

just 167 lysines in internal α-helical positions. As noted by Blaber et al., this data is clearly statistically insufficient. For this reason, Blaber et al. estimated the conformational entropy of lysine and other long side chains using just the first two side-chain torsions. Pickett and Sternberg ***(7)*** and Abagyan and Totrov ***(6)***, in their derivations of empirical scales of side-chain entropies in protein folding, encountered the same problem. They also used just the first two side-chain torsions for long side chains, and added a residue-specific term estimated by making the assumption that the third and higher torsions were equally distributed among available states.

Modeling conformational states of unfolded states leads to the major assumption of this approach. One assumes that the distribution of conformations observed in all structures (or some subset of structures) within the dataset approximates that of the unfolded states of proteins. This is probably a reasonable assumption, although there is little evidence to support (or refute) it.

3.3. Monte Carlo Computer Simulations

Monte Carlo computer simulations of atomic/molecular systems were first introduced by Metropolis et al. ***(14)*** nearly 50 yr ago. This method of simulating molecular systems has proven to be very effi-

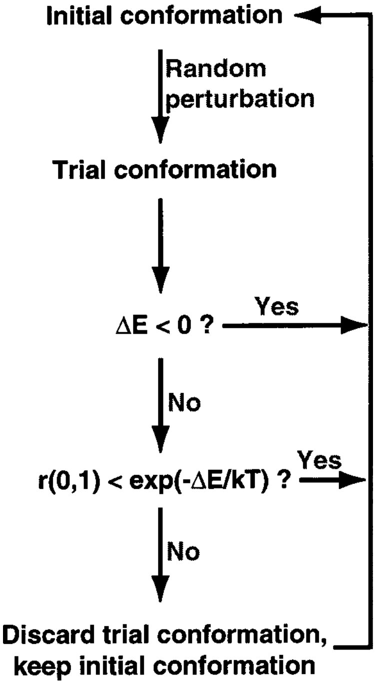


Fig. 3. The Metropolis ***(14)*** algorithm for Monte Carlo computer simulations.

cient for exploring allowable regions of conformational space for molecules with more than a handful of degrees of freedom. The basic Metropolis algorithm is outlined in **Fig. 3**. The algorithm, and numerous sampling schemes, are discussed in detail elsewhere (*see* **ref.** ***15***), and will not be covered in great detail here.

A basic outline of the method is as follows: one starts with the molecular system—for example, a peptide—in an initial conformation. This conformation is then perturbed in some manner. For example, a randomly chosen bond may be rotated by a random amount in order to generate a new conformation. The difference in energy between the new and initial conformations, ΔE, is calculated. If the new conformation is of lower energy ($\Delta E < 0$), it is retained, and becomes the new initial conformation. If, on the other hand, the new conformation is of higher energy ($\Delta E > 0$), it is kept only if a randomly generated number between zero and one, r(0,1), satisfies the following

$$r(0,1) < e^{-\Delta E / RT}. \tag{6}$$

Otherwise, the newly generated conformation is discarded. This scheme is repeated in an iterative manner many thousands, even millions, of times (**Fig. 3**). The resulting set of conformations will approximate the Boltzmann distribution and can be partitioned into states. The conformational entropy of the system can then be estimated using **Eq. 1**. Errors in the conformational entropy can be estimated using standard methods such as those outlined by Smith and Wells ***(16)***.

Monte Carlo simulations have been employed to generate conformational distributions for estimating the conformational entropy lost by side chains folded into α-helices ***(2,3)*** and for generating a scale of side-chain conformational entropy in unfolded states ***(17)***. In principle, one could use Monte Carlo computer simulations to estimate conformational entropy for large systems, potentially even whole proteins. In practice, computer power and error estimates limit such calculations to the level of a few residues or less.

The major advantage of using Monte Carlo computer simulations compared to a dataset of protein structures is that it is possible to generate as much data as needed to adequately approximate the conformational distribution in question. For example, if one is interested in the conformational distributions of lysine side chains, it is possible to generate many thousands of lysine conformations relatively rapidly using an efficient Monte Carlo sampling scheme. Methods for determining what is “enough” data from a computer simulation are described by Kolafa ***(18)***.

There are three potential problems with this approach: the energy potential or force field employed, treatment of solvent, and modeling the unfolded state. The energy potential employed in a Monte Carlo simulation essentially determines the conformational distribution that is produced. One hopes that the potential being used reproduces the conformational behavior of a real peptide or protein. All available potentials are approximations in one way or another.

However, modern potentials such as AMBER94 ***(19)*** or class II force fields (CFF) ***(20)*** appear to be capable of reproducing many of the conformational properties of proteins. A comparison of generated distributions with conformations observed in protein structures can help validate the results obtained.

Treatment of solvent is problematic in Monte Carlo computer simulations. Sampling schemes often employ potentially large rotations about bonds, resulting in large movements of parts of the molecule being simulated. If an explicit solvation model were to be employed (i.e., solvent molecules included), such large movements would invariably result in steric clashes between the molecule and solvent, leading to rejection of the newly generated conformation under the Metropolis scheme (*see* **Fig. 3**). There are sophisticated sampling schemes designed to overcome this problem; yet these are beyond the scope of this chapter. More commonly, solvent is treated either as a dielectric continuum, or with an implicit solvation model such as generalized born/surface area (GB/SA) ***(21)***. Implicit solvation models have advanced rapidly in recent years, and are approaching the level of accuracy previously achieved only with explicit solvation models. One can expect reasonable conformational distributions employing such an implicit solvation model with a modern energy potential.

As with distributions derived from a protein structure dataset, modeling the unfolded states of proteins is difficult. With the Monte Carlo approach, the unfolded states are generally modeled using something as simple as a tripeptide model. For example, in studies of side-chain conformational entropy, the unfolded states have been modeled using Ace-Ala-Xaa-Ala-NMe, where Xaa was the residue under consideration ***(2,3,17)***. It is debatable that such a simple model is an accurate representation of the conformational properties of unfolded proteins. However, it is difficult to adequately sample the available conformational space of larger, more complex models.

3.4. Molecular Dynamics Computer Simulations

Molecular dynamics (MD) simulations are not a good method of generating conformational distributions for estimating conformational entropy, because rotations about bonds occur on much the same time scale (picoseconds) as the average simulation (100–1000 picoseconds). MD simulations are simply not an efficient method for searching all available conformational space.

3.5. Exhaustive Conformational Searches

The exhaustive conformational search is another computer-based method that has been used to estimate conformational entropy. Rather than generating a distribution of conformations, this approach generates an energy distribution which can be used in conjunction with **Eqs. 4** and **5** to obtain entropy estimates. This approach has been used to estimate side-chain conformational entropy loss for side chains folded into an α-helix ***(4)***, the entropy loss associated with interactions between pairs of side chains in helices ***(8)***, and the difference in backbone conformational entropy between alanine and glycine ***(9)***.

The basic algorithm for a conformational search is illustrated in **Fig. 4** for a leucine side chain. This example is used to explain the workings of this method. In leucine, there are two rotatable side chain dihedrals—χ_1 and χ_2 (ignoring hydrogens). In a conformational search of a leucine side chain, one starts by rotating χ_1 by a fixed increment (often 5° or 10°). The energy of that conformation is then calculated—often after minimizing the energy while keeping the side-chain dihedrals fixed. The second dihedral, χ_2, is then rotated by the fixed increment and the energy is calculated. One keeps rotating about χ_2, calculating the energy at each step, until 360° have been traversed. The first dihedral, χ_1, is then rotated by the fixed increment, χ_2 rotated through 360°, and so on (*see* **Fig. 4**). Thus, for each incremental rotation of χ_1, χ_2 is rotated, in increments through 360°. This is continued until χ_1 has traversed 360°. The resulting set of energies describe the energy surface for the leucine side chain. Extending this method to other side chains, or to the backbone, is a straightforward process.

The caveats that apply to the Monte Carlo approach—accuracy of the energy potential, treatment of solvent and modeling unfolded

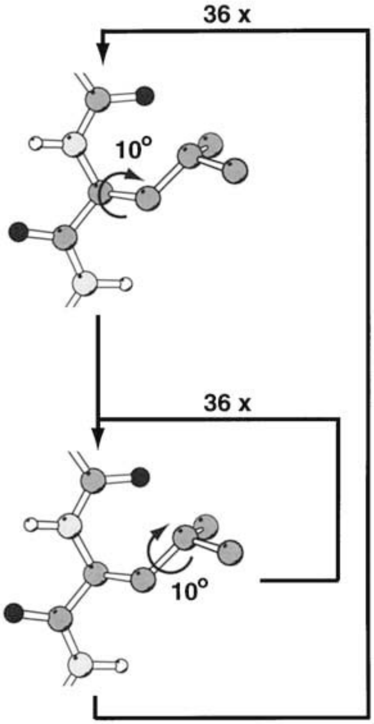


Fig. 4. Exhaustive conformational search of all conformations available to a leucine side chain.

states—all apply to exhaustive conformational searches. The modeling of unfolded states is particularly problematic, since this method becomes extremely intensive computationally for systems involving more than four or five rotatable dihedrals. For such systems, the Monte Carlo approach is generally more efficient. One possible approach—that taken by Lee et al. ***(4)***—is to use entropy estimates for the unfolded states generated using Monte Carlo simulations, and to generate entropy estimates in the folded state using conformational search methods.

3.6. Other Methods

There are other methods of estimating conformational entropy ***(4,22)***. For example, Sternberg and Chickos ***(22)*** estimated side-

Table 2
Software Available for Performing Biomolecular Monte Carlo Computer Simulations and/or Exhaustive Conformational Searches[a]

Software	Provider	WWW Address
InsightII/Discover	Molecular Simulations, Inc.	www.msi.com
Cerius^2	Molecular Simulations, Inc.	www.msi.com
MacroModel	Schrödinger, Inc.	www.schrodinger.com
HyperChem	Hypercube, Inc.	www.hyper.com
Spartan	Wavefunction, Inc.	www.wavefun.com
BOSS	Dr. W.L. Jorgensen, Yale University	zarbi.chem.yale.edu
Fantom	Dr. W. Braun, University of Texas, Medical Branch	www.scsb.utmb.edu/fantom/fm_home.html

[a] Note: The software is listed in random order, and the author does not endorse any particular software package or company. The author apologizes to those whose software is not included above. Any exclusions are unintentional.

chain conformational entropy using experimentally determined entropies of fusion for small organic compounds. The results of this approach are in excellent agreement with data obtained using the methods discussed in **Subheadings 3.2.–3.5.** Such alternative methods are not commonly used, and as such are not discussed further here.

4. Software

A list of some of the available software capable of performing Monte Carlo simulations and/or conformational searches is provided in **Table 2**. This list is not intended to be comprehensive.

References

1. Aurora, R., Creamer, T. P., Srinivasan, R., and Rose, G. D. (1997) Local interactions in protein folding: lessons from the α-helix. *J. Biol. Chem.* **272,** 1413–1416.

2. Creamer, T. P. and Rose, G. D. (1992) Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. USA* **89,** 5937–5941.
3. Creamer, T. P. and Rose, G. D. (1994) α-Helix-forming propensities in peptides and proteins. *Proteins* **19,** 85–97.
4. Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994) Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins* **20,** 68–84.
5. Blaber, M., Zhang, X. J., Lindstrom, J. D., Pepoit, S. D., Baase, W. A., and Matthews, B. W. (1994) Determination of α-helix propensity within the context of a folded protein: sites 44 and 131 in bacteriophage T4 lysozyme. *J. Mol. Biol.* **235,** 600–624.
6. Abagyan, R. and Totrov, M. (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235,** 983–1002.
7. Pickett, S. D. and Sternberg, M. J. E. (1993) Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* **231,** 825–839.
8. Creamer, T. P. and Rose, G. D. (1995) Interactions between hydrophobic side chains within α-helices. *Protein Sci.* **4,** 1305–1314.
9. D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25,** 143–156.
10. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. R., et al. (1997) The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112,** 535–542.
11. Hobohm, U., Scharf, M., Schneider, R., and Sander, C. (1992) Selection of representative protein data sets. *Protein Sci.* **1,** 409–417.
12. Hobohm, U. and Sander, C. (1994) Enlarged representative set of protein structures. *Protein Sci.* **3,** 522–524.
13. Koehl, P. and Delarue, M. (1994) Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239,** 249–275.
14. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21,** 1087–1092.

15. Allen, M. P. and Tildesley, D. J. (1992) "Computer Simulation of Liquids," Oxford University Press, Oxford, UK, pp. 1–385.
16. Smith, E. B. and Wells, B. H. (1984) Estimating errors in molecular simulation calculations. *Mol. Phys.* **52,** 701–704.
17. Creamer, T. P. and Rose, G. D. (1995) Simple force field for study of peptide and protein conformational properties. *Meth. Enzymol.* **259,** 576–589.
18. Kolafa, J. (1986) Autocorrelations and subseries averages in Monte Carlo simulations. *Mol. Phys.* **59,** 1035–1042.
19. Cornell, W. D., Cieplak, P., Bayley, C. I., Gould, I. R., Merz, Jr., K. M., Ferguson, D. M., et al. (1995) A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117,** 5179–5197.
20. Maple, J. R., Hwang, M. J., Stockfish, T. P., Dinur, U., Waldman, M., Ewig, C. S., et al. (1994) Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comp. Chem.* **15,** 162–182.
21. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112,** 6127–6129.
22. Sternberg, M. J. E. and Chickos, J. S. (1994) Protein side-chain conformational entropy derived from fusion data—comparison with other empirical scales. *Protein Eng.* **7,** 149–155.

7

Turn Scanning

Experimental and Theoretical Approaches to the Role of Turns

Carl Frieden, Enoch S. Huang, and Jay W. Ponder

1. Introduction

The mechanism that enables a protein to fold to its correct three-dimensional conformation—given only the information available from its sequence—has been a matter of intense investigation since Anfinsen's classical experiment on the refolding of ribonuclease almost 40 years ago ***(1)***. The importance of turn sequences in the kinetic and thermodynamic aspects of protein folding remains a controversial point, although their role—especially for β-turns—has been regarded as significant (i.e., ***2–10***). What role, if any, do turn sequences play in determining the native structures of proteins? For the experimentalist, this question is usually rephrased as: to what extent can substitution, insertion, and deletion be tolerated at turn positions before kinetic and/or thermodynamic properties are affected? It is indeed surprising how few experimental studies have been performed to carefully investigate the role of turns. To the theoretician, this question of the importance of turns revolves around the ability of computational studies to attempt to determine the propensity of isolated small peptides, as well as larger struc-

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

tures, to form turns. While much research has been conducted in this field, even here the answers remain uncertain.

In this chapter, we draw on both experimental and theoretical results in an attempt to answer the question. Part of the debate in the experimental literature concerns the issue of whether turns play an active or passive role in protein folding. If we accept the theory that isolated turn sequences may be intrinsically unstable, it may be tempting to conclude that turns will play only a passive role in protein folding. While this view may have value for very small peptides, it can be overly simplistic for at least two reasons. First, there is evidence that sequence-specific intraturn side-chain interactions can stabilize turn structures. Second, early, nonspecific stages of protein folding—such as a generalized hydrophobic collapse phase—could supply additional constraints that might allow turn sequences to act as nucleation sites for a more ordered structure. For example, hydrophobic interactions within a folding β-hairpin may restrict the range of extended structures available to the included turn and disrupt the turn-water entropic effects favoring extended structures.

Here we present the use of turn scanning as an experimental method for assessing the role of turns in the folding process. Turn scanning—in analogy to methods such as alanine scanning—is the systematic replacement of residues in a turn followed by an examination of the consequences of the change in terms of rates of folding or unfolding, stability to denaturing conditions, and structural changes using various biophysical techniques. An obvious candidate for replacement would be glycine. Replacing this residue in a turn may lead to a change in the turn type. Similarly, replacing a residue with glycine may lead to greater flexibility within the turn. Another obvious candidate is proline, involving either an insertion or a replacement. Changes should certainly not be limited to these two residues. Substitutions between hydrophobic and hydrophilic residues, or between residues with a large and small side chains, are equally valuable. For example, replacing a leucine in one of the turns of the fatty-acid-binding protein was found to lead to a very unstable protein that folded rapidly to an intermediate state ***(11)***.

However, one must be careful with the results. The permissive substitution of native-turn residues does not necessarily imply a passive role for the turn. The native-turn sequence may in fact actively drive or stabilize hairpin formation more than other segments of similar length, but in the absence of the native sequence, other interatomic forces may be sufficiently strong to drive folding, making the mutated sequence a turn by default. This point is made eloquently in the work of DeGrado and colleagues ***(12)***. By the same logic, extreme intolerance to substitution does not necessarily mean that the native sequence plays an active role in folding. For example, the substitution of another residue for glycine in a turn may abolish folding because of the steric clashes resulting from the presence of a beta-carbon. Does the glycine actively direct turn formation, or is its presence merely required, as it is the only amino acid that does not destabilize the requisite local conformation? The point is to caution that equilibrium and kinetic experiments should be coupled with other biophysical techniques to resolve the issue of the importance of turn structure in the folding process.

2. Kinetic Evidence for the Importance of Turn Formation in the Folding Process

2.1. Experimental Results from Model Systems

Researchers have generally chosen peptides or small proteins as model systems. In either case, the rationale is that they exhibit simple two-state behavior—i.e., the protein is either in an unfolded or folded state with no populated intermediate state. While in reality, many systems are not two-state—especially when investigated using site-directed mutagenesis to perturb the folding process—the results from these model systems provide evidence for the importance of turn formation in protein folding.

2.2. Peptide Model Systems

A logical approach to measure the rate of turn formation is to utilize peptides with a natural tendency to form β-turns, and then perturb the system by using different amino acids in specific positions. Several investigators have done this. De Alba et al. ***(13)*** have constructed model linear peptides in which the only difference was in the putative turn region. Their results clearly demonstrated that the turn residue sequence determined the turn conformation, and thereby other features such as interstrand pairing and hydrogen bonding. Munoz et al. ***(14)*** have studied the kinetics of one peptide (16 aa) that reflected the β-hairpin from protein G B1 using laser temperature-jump and found that folding of the hairpin occurs in 6 μs at room temperature. This paper is followed by a statistical mechanical model for β-hairpin kinetics ***(15)*** in which the authors attempt to predict the kinetic properties of other similar peptides (*see also* Chapter 8). The results of the work of Serrano and colleagues ***(16–19)*** are also particularly pertinent. Blanco and Serrano—using circular dichroism and NMR (nuclear magnetic resonance)—have examined the solution structure of the isolated fragments 1–20 (β-hairpin), 21–40 (α-helix), and 41–56 (β-hairpin), corresponding to all the secondary-structure elements of the protein G B1 domain. Turn-like folded structures were detected in water, although they were poorly populated ***(16)***. Ramirez-Alvarado et al. ***(17)*** have designed a 12-residue model peptide that folded into a monomeric β-sheet. Substitution of strand residues by alanine led to the loss of the hairpin structure, demonstrating the importance of side-chain interactions between the strands. Mutations in the turn of the same peptide ***(18)*** showed that changing the central residue of a type I′ β-turn affected stability. These and other experiments from this laboratory ***(7)*** led Blanco et al. ***(19)*** to emphasize the importance of residues within and near the turn in stabilizing β-hairpin structures. It is interesting that umbrella sampling with the CHARMM force field and ACE implicit solvation model has been used to estimate the stability of the peptide examined by Ramirez-Alvarado et al. ***(18)***. These calculations ***(20)*** predicted a 38% content of hairpin structure at room temperature with less than 1% helical content, and the remainder in various unfolded states. There appears to be a much lower barrier for this transition between the hairpin and unfolded states than for the corresponding helix-coil transition.

There is little doubt that changing residues within turns of model peptides affects stability and conformation, as many studies have now shown. However, only a few of these studies have been directed toward the effect of the change on the kinetic properties. This may be a consequence of the fact that very rapid techniques (i.e., µs or shorter) are not generally available. Clearly, such experiments are essential for evaluating the role of turns in the kinetics of folding. Theoretical results may thus become complementary to experimental results—as the former tries to extend simulations to longer times, while the latter attempts to determine events that occur at early times.

2.3. Protein Model Systems

The intestinal fatty-acid binding protein is an excellent model for folding studies because it is almost entirely β-sheet, and 7 of the 8 turns between β-strands include a glycine residue. Kim and Frieden ***(21)*** carefully examined the effect of mutations in turns on the stability and refolding of this protein by mutating each of these glycines to valine. In some cases, the mutation had little or no effect on folding or stability, while in other cases the effect was dramatic. Specifically, mutations in type II turns—while very destabilizing—did not markedly affect the rate of structure formation, while the mutation of a glycine to valine in the turn between the two last β-strands of intestinal fatty-acid binding protein (IFABP) slowed the rate of folding by at least 100-fold. Furthermore, the refolding rate constant for this mutant was observed to be essentially independent of the final denaturant concentration ***(21)***. Random mutagenesis in another turn (between the D and E strands) led to the conclusion that a leucine residue was critical for formation of the final stable structure and perhaps for an initial nucleation step as well ***(11)***. These experiments represent one of the first attempts to critically evaluate the role of different turns in the overall folding process.

While almost all of the turn mutants of the fatty-acid binding protein examined were able to fold and bind fatty acid, a very different observation was made for plastocyanin. Ybe and Hecht ***(22)*** found that the vast majority (92 of 98) of four-residue turn mutants of this protein apparently could not fold into the native β-barrel topology. Results of Gu et al. ***(23)*** also highlighted severe sequence constraints in two β-turns of peptostreptococcal protein L. Because their mutants were presented on a phage display system, enumeration and sequencing of all the mutants was not possible. Nevertheless, biopanning against the immunoglobulin G (IgG) binding activity of protein L showed that less than 0.1% of the phage from the two libraries (one for each turn that was mutagenized) were recovered.

Schonbrunner et al. ***(9)*** have presented evidence that hairpins may serve as initiation sites for β-sheet formation. These authors used the small (74 aa) β-sheet protein tendaminstat that contains two disulfide bonds, one of which connects the ends of a β-hairpin. Using site-directed mutagenesis of one of these cysteines, they found that unfolding is greatly enhanced, but with a greatly reduced effect on the refolding rate. Using a Fersht analysis ($\alpha > 1$), they concluded—based on thermodynamic and kinetic arguments—that formation of the hairpin turn is the rate-limiting step in the refolding process.

Some studies suggest that turns connecting α-helices are rather robust with respect to mutation. In the early work by Hecht and colleagues ***(24)***, a three-residue turn in the four-helix bundle cytochrome b-562 was randomized to 31 unique sequences, and all the mutants folded correctly. Subsequent experiments on the four-helical Rop dimer provide corroborating evidence. In one study, Regan and colleagues ***(25)*** changed one of the two residues (Asp30) involved in the tight turn between the two helices of the monomer into every other amino acid without misfolding the protein. In another study, Cesareni and colleagues ***(26)*** mutated three residues semirandomly in the turn region of the Rop monomer (residues 30 to 32), and found only three mutants out of a possible 380 that failed to fold correctly. Yet substitution of a proline for alanine in the turn caused drastic changes ***(27)***. MacBeath et al. ***(28)*** conducted a highly comprehensive randomization of an interhelical turn in Escherichia coli chorismate mutase—also a four-helix bundle. The authors reported that more than 63% of tripeptide sequences can functionally substitute for the native-turn sequence in this enzyme.

Nagi et al. ***(29)***, examining the folding pathway of the four-helix bundle protein, Rop, replaced the wild-type two-residue loop between helical regions with a series of polyglycine linkers of increasing length. They then investigated the kinetics of unfolding and refolding, and observed a large dependence of rates as the loop length increased from 2 to 4, but little dependence beyond that. They suggested that the results were consistent with lowering the energy barrier of the transition state, and that this type of experiment can test potential folding models. Similar experiments on changing loop lengths have been reported by Viguera and Serrano ***(30)*** and by Ladurner and Fersht ***(31)***. The results of this type of experiment may reveal the importance of turn formation relative to interactions between structural units on either side of the turn.

Currently, there is a growing database that describes the altered stability of proteins with mutated turns. The recurrent theme is that mutations in loops generally reduce the thermodynamic stability of the folded state ***(11,12,23,32,33)*** although exceptions do exist ***(25)***. Consequently, misfolded mutants may correspond to those whose altered turn residues—if forced to assume the native conformation—would exact a thermodynamic penalty exceeding that of the scaffold's overall stability ***(12)***. This line of reasoning thus offers two factors for the differential sensitivity towards turn mutation: 1) the inherent stability of the scaffold, and 2) the energetic demands upon the turn in question.

3. How to Relate Changes in Kinetic Parameters to the Importance of Turn Formation in the Folding Process

Much of the interpretation of the effect of mutations on rate constants of unfolding or refolding in the presence of chemical denaturants has been based on a simple two-state model, N ⇔ U. Jackson and Fersht ***(34)*** have discussed chevron plots (i.e., plots of the log of the relaxation rate or rate constant as a function of the denaturant concentration) for the two-state model. The results of mutagenesis of a wild-type protein can affect these plots in at least three ways: an equilibrium mutation in which both folding and unfolding rate constants are affected equally with no change in stability; a kinetic mutation in which one of the rate constants is changed, along with a change in stability, but the other is not; or a mixed equilibrium-kinetic effect. Fersht and colleagues ***(35)*** have developed the use of the quantity ϕ (=$\Delta\Delta G_{I-D}/\Delta\Delta G_{N-D}$) to determine the mutations effect on processes before or after the transition state for a folding pathway that includes an intermediate. In this calculation, this ratio is the change in free energy on mutation relative to the overall free energy of folding. If $\phi = 1$ or 0, the interpretation is reasonably straightforward, reflecting whether the transition state is perturbed by mutation to the same extent as the denatured state ($\phi = 1$) or the native state ($\phi = 0$). Unfortunately, the actual interpretation is more difficult. First, many systems show more than one rate in the folding process, suggesting the presence of intermediates. Frequently, there is a burst too rapid to measure by normal stopped-flow techniques. Second, one must consider what the rates are actually measuring. For example, using changes in fluorescence (or A or circular dichroism, the most common methods of following folding or unfolding), it is usually unclear what the contribution of intermediate states to the observed change actually is. Third, there is the issue of whether observed phases represent intermediates or off-pathway forms. Fourth, the log of the rate constant is not always a linear function of the denaturant concentration (i.e., the chevron plot is nonlinear). Certainly, proline isomerization will affect this linearity, but such observations can be made even for a protein that does not contain proline ***(21)***. The issue of the nonlinearity in chevron plots unrelated to proline isomerization has been discussed by other authors ***(36,37)***. From this perspective, it is easy to see why investigators want to use model peptides. For proteins, however, all is not lost. Despite these complexities, some generalizations are possible. For example, if the amplitude of a refolding burst phase is decreased by the mutation, it is likely an early step that is affected. Similarly, if the burst phase is unaffected, a late step may be responsible. But there are caveats for even this simple interpretation. Thus, for comparison of a mutant to wild-type, one must know the midpoint and the degree of cooperativity of the equilibrium denaturation curve. A loss in cooperativity (the *m* value of the 6-term equation that describes equilibrium denaturation measurements ***[38]***) suggests increased concentrations of one or more intermediates as a result of the increase in the rate constant for the formation (or more likely, a decrease in that of unfolding) of that intermediate. At the same time, however, a decrease in cooperativity makes it difficult to know what to compare between the mutant and the wild-type. For a highly cooperative denaturation, one can set the difference between the midpoint of the denaturation curve and the final denaturant concentration to be equal for comparative purposes. For noncooperative denaturation, there is no simple way to compare the rate constants for folding or unfolding in the mutant to those of the wild-type protein.

Another criterion is the slope of the chevron plot—i.e., how the rate constant changes as a function of the final denaturant concentration (ignoring effects of proline isomerization and off-pathway forms). The rate constant as a function of denaturant concentration is frequently defined as:

$$\ln k = \ln k_0 -/+ (m/RT)[D] \tag{1}$$

where k_0 is the rate constant in the absence of denaturant, *m* is related to extent of cooperativity in the equilibrium denaturation curve as mentioned here and above, *D* is the denaturant concentration, and the –/+ indicates whether the rate constant examined is for folding or unfolding. Thus, a plot of ln *k* as a function of denaturant concentration (at constant temperature) should have a slope directly proportional to *–m* (or *m* in the unfolding direction). If the equilibrium denaturation curve is asymmetric, the value of *m* may be different in the folding and unfolding directions. Thus, in theory, the slope and the cooperativity are directly correlated. It is important to remember, however, that these rate constants are apparent values, and include the microscopic rate constants of all the steps in the process. For a simple reversible process:

$$N \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} U \tag{2}$$

the observed rate constant is $k_1 + k_{-1}$. For the process:

$$N \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} I \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} U \tag{3}$$

two apparent rate constants, λ_1 and λ_2, may be observed. In this case the observed rates are (approximately) related to the microscopic rate constants by the relationships $\lambda_1 = k_1+k_{-1}+k_2+k_{-2}$ and $\lambda_2 = [k_1(k_{-2}+k_2) + k_{-1}k_{-2}]/(k_1+k_{-1}+k_2+k_{-2})$, respectively. For extreme dilutions of the denaturant (in the refolding direction) these may simplify to k_1+k_2 and $[k_1k_2]/(k_1+k_2)$ and if $k_2 << k_1$, then the apparent constants become the microscopic constants. It is not surprising that experimental values frequently deviate from theory for systems that exhibit several phases.

4. Rate Measurements: Experimental Procedures

Whether using fluorescence, absorbance, circular dichroism, or other similar techniques, most standard stopped-flow mixing devices have dead times on the order of 1–2 ms. Most proteins appear to form a collapsed state well within the first ms (the "burst" phase). Slower steps seem to be associated with the stabilization of the final structure. For example, based on stopped-flow ^{19}F NMR data, the slow steps in the refolding of *E. coli* dihydrofolate reductase are stabilization of side chains near the protein surface ***(39–41)***. Similarly, mutations in one turn of the fatty-acid-binding protein allow formation of an intermediate without the final stabilizing step ***(11)***. Thus, combining these stopped-flow experiments with turn scanning may help define the role of turn formation in obtaining the final stable protein structure.

It is clearly important to explore the effect of mutations on early events. We have already discussed changes in the amplitude of a burst phase. A better method may be the use of faster mixing methods to determine changes in apparent rate constants that are hidden in the burst.

With the advent of experimental techniques that can measure half-times of folding processes faster than the usual 1-ms stopped-flow dead time, it should be possible to see whether mutations affect early steps in folding. The use of faster techniques has recently been reviewed by Callender et al. ***(42)*** and by Plaxco and Dobson ***(43,44)***. These new approaches for monitoring faster events use certain techniques with time resolution as fast as picoseconds. Perhaps the simplest of these is the continuous-flow method using capillary tubes, which allows mixing within 50–100 µs ***(45)***. Several studies have now used this method ***(46,47)***. Other techniques include laser-induced temperature-jump, laser photolysis, optical electron transfer, fluorescence correlation spectroscopy, or energy transfer and NMR.

Oas and colleagues ***(48–51)*** have used dynamic NMR to investigate simple two-state systems by extrapolating observed rate constants back to zero denaturant concentrations to obtain values for rapid steps in the µs range for the refolding process. Native-state amide proton exchange has been used by Arrington and Robertson ***(52)*** with ovomucoid third domain to determine rates of folding and unfolding in the µs range. Gilmanshin et al. ***(53)*** have used temperature-jump (10-ns, laser-induced) and time-resolved IR spectroscopy in native apomyoglobin, and observed two relaxation phases with relaxation times of 48 ns and 132 µs. They attributed these to formation of substructures primarily related to helices. The slower of the two phases is attributed to tertiary interactions and desolvation of the helix, but the data are consistent with substructures guiding the folding process and with these substructures being independently formed. Ballew et al. ***(54)*** also used temperature-jump with this protein, and noted that the earliest steps during folding to a compact state were complete within 20 µs. Nolting et al. ***(55)*** have also used temperature-jump methods on the cold denatured mutant of barstar to obtain relaxation rates in the µs range. In a later article, Nolting et al. ***(56)*** examined the folding pathway over a time range from µs to s again using fluorescence changes as measured by temperature-jump techniques to examine the early times. Their data are consistent with a nucleation-condensation mechanism.

For the most part, however, these techniques have not been applied to systems in which there are mutations, especially where those mutations are in turns. As these methods become more available, such experiments will hopefully be done.

5. Experimental Approach to Selecting Mutations

Certainly there are no general rules for deciding what mutations to make in a given turn. As discussed in this chapter, from both an experimental and theoretical point of view, substitutions between glycine and other residues or between proline and other residues are the most obvious choices. Hutchinson and Thornton ***(57)*** have presented extensive data for sequence preferences in different β-turn types that could be useful in deciding what mutation to make. Aside from glycine and proline changes, one should consider other steric or hydrogen-bonding effects. Substitution of a β-branched chain amino acid for one that has φ, ψ angles in a nonallowed region will certainly change the turn type, as will almost any substitution of a glycine. Substitutions between hydrophobic and hydrophilic residues may alter the solvent accessibility of the turn. While positions 2 and 3 of a β-turn are obvious candidates for substitution, residues adjacent to the turn—especially those that may affect packing within the turn—should be considered. Random mutagenesis of the two residues in the turn could be useful, but with the wide range of possible mutant proteins, a screen for function may be necessary.

6. Computational Studies: Theory for Turn Folding and Structure

Over the past decade, molecular dynamics and other simulation techniques have been used to explore the structural propensities and conformational dynamics of a variety of short turn-like peptide sequences. Advances in force-field accuracy and available computer power now allow simulation to serve as a viable complement of experiment. Indeed, because of the all-atom level of detail provided by simulation, it could eventually become a preferred method for interpreting and correlating more ambiguous structural results. For example, reliable theoretical estimation of conformational population densities would greatly aid the interpretation of ambiguous NMR data for flexible peptides. Steps toward this ultimate goal are already underway.

Molecular dynamics simulation of turn peptides with explicit representation of solvent is still limited to the ns to ms time-scale. Fortunately, many of these sequences adopt relatively self-contained structures within full proteins, and are usually devoid of large, energetic barriers to conformational interconversion. Thus, even ns-scale MD simulations can expose some of the structural variability and dynamic behavior of turns. Two early reviews of peptide simulations ***(58,59)*** provide an excellent survey of this field through 1993. Here, we present an update and attempt to infer some general principles regarding the behavior of turns in protein structure.

In early work, Tobias et al. ***(60)*** used umbrella sampling methods to explore the free-energy surfaces of the blocked dipeptides Ac-Ala-Ala-NMe and Ac-Pro-Ala-NMe in aqueous solution. Based on calculated equilibrium constants for unfolding of type I turn structure, they found that the extended conformations were favored for both peptides by 5 to 10 kcal/mol. Additional studies with glycine replacing alanine in the second position suggested that glycine destabilizes the extended states, but the manifold of type I turn structures still lies 3 kcal/mol higher in free energy ***(58)***. Further simulations of the stability of type II turns suggest that Ac-Ala-Gly-NMe may be marginally stable as a type II turn. All other sequences again favor the extended form. The Ala-Gly result is corroborated by independent simulations from a second group ***(61)***.

Molecular orbital results at the HF/6-31G* level also predicted that the extended forms of capped Gly-Gly and Ala-Ala are enthalpically favored over type I and II turns by values in the 1–9 kcal/mol range ***(62)***. Type I′ and II′ structures are even less favorable for these sequences. As expected, inclusion of electron correlation at the MP2 level favors the hydrogen-bonded turn forms. However, the authors claim that correlation energy effects and intramolecular entropic terms compensate so that the HF energies are a fair approximation to the true enthalpy differences. *Ab initio* MO calculations coupled with the PCM solvation model ***(62)***, MD with explicit solvent, and finite-difference Poisson-Boltzmann calculations with the MEAD package ***(63)*** all agree that solvation effects favor type I over type II turns for all combinations of amino acids at positions 2 and 3 of a β-turn. Decomposition of the free-energy difference between extended and turn conformations indicates that peptide-water entropy is the major term favoring extended structures. Finally, Monte Carlo sampling of a variety of blocked dipeptides using the CHARMM19 force field plus Poisson-Boltzmann solvation ***(64)*** again finds the extended form to be favored over turn types I, I′, II and II′ by 1.6 to 7.7 kcal/mol depending on the sequence. However, with the exception of Gly-Gly, the most stable turn type for each sequence was found to be the one most commonly observed in protein structures. The Gly-Gly occurs frequently in type I′ turns of β-hairpins where the strand twist is apparently able to overcome local energetic effects.

Thus, both force-field and quantum mechanical results for these small peptides lead to the conclusion that reverse turns are intrinsically unstable in the absence of side-chain interactions and the tertiary protein environment.

Do turn types interconvert? Reaction path calculations on a capped Ala tripeptide as a model reverse-turn sequence ***(65)*** find only low barriers between extended and turn-like conformations. Interconversion pathways tend to involve sequential rotation about single backbone dihedral angles. This is in agreement with further path calculations performed using the CHARMM force field on an analog of the Ala tetrapeptide ***(66)***. Work with other force fields (*see* **ref. *66a***), semiempirical quantum mechanics and detailed analysis of protein crystal structures ***(67)*** also suggest low barriers for interconversion, but find pathways involving concerted dihedral angle rotations. For example, the MOPAC results of Gunasekaran et al. ***(67)*** assign an enthalpic barrier of only 3 kcal/mol to the concerted peptide bond flip needed to convert type II and type I β-turns. In summary, theory suggests interconversion between turn types, and between turn and nonturn structure, is energetically accessible, and may play an important role in peptide and protein dynamics.

7. Importance of Side-Chain Packing: The Type VI Turn

We indicated earlier that logical candidates for mutations in turns are either removal or insertion of glycine or proline. We now examine why, based on computational studies, the incorporation of a proline residue into position 3 of a turn may affect stability and/or kinetic properties. A well-studied example of a stable turn peptide containing proline is the Ser-Tyr-Pro-Phe-Asp-Val sequence, which adopts a type VI turn structure. NMR analysis finds a 70% population of type VIa and VIb turns in aqueous solution ***(68)***. A theoretical study of this same peptide using locally enhanced sampling (LES) methods reaches generally similar conclusions ***(69)***. These workers also find a type VI turn structure to be the preferred conformation. Their simulations suggest a possible kinetic pathway for the folding in this peptide family. First, electrostatic interactions involving the N- and C-termini and aspartate residue tend to enforce an end-to-end distance consistent with turn formation. Next, the backbone angles of the peptide rearrange in concert with a generalized hydrophobic collapse involving side chains. Finally, the hydrophobic residues pack into the characteristic type VI turn. This work also uncovers some potential "trapped" states containing *trans*-Pro residues. Molecules with the *trans*-linkage remain disordered with large end-to-end distances, and are not stabilized by hydrophobic interactions.

An independent study of the related Ala-Tyr-*cis*-Pro-Tyr-Asp-NMe sequence uncovered similar hairpin stabilizing factors. In contrast to the simulation results described for generic sequences, this peptide is stable as a turn structure during lengthy MD simulations.

A 20-ns trajectory starting from an NMR-derived type VIa structure makes three transitions between type VIa and VIb turns, but does not explore extended conformations ***(70)***. A further simulation, starting from an unfolded structure, folds to a compact turn structure within 2.7 ns. Finally, a 4.7-ns simulation starting from a nontype VI *cis*-Pro conformation is trapped in an unfolded state, which is stable over the course of the trajectory. Subsequently, a potential of mean force (PMF) constructed from the AMBER/OPLS energy function plus macroscopic solvation terms was used to select low-energy conformers from a random search ***(71)***. The PMF selected a structure similar to the NMR model as the lowest energy conformer. Inclusion of entropy estimates led to a final $\Delta G_{folding}$ in the range –1.0 to –2.1 kcal/mol. These values are somewhat more favorable to turn formation than the NMR studies, perhaps because of incomplete sampling of the manifold of unfolded states.

Two factors indicate why is this particular type VI turn sequence is stable in a compact turn-like conformation. First, the proline residue at turn position 3 favors a *cis*-dihedral angle at the preceding peptide bond. The relative advantage is approx 2.6 kcal/mol over the *trans*-peptide bond favored for other amino acids ***(72)***. Equally or more important for this sequence is the packing of the two aromatic tyrosine residues against the proline ring in the type VI turn structure. In fact, the PMF calculations indicate that intrapeptide enthalpic interactions comprise the major term favoring the folded form, while polar solvation effects favor extended conformations.

Other theoretical work from Pettitt’s group ***(73)*** examined the conformation of the tetrapeptide tuftsin, Thr-Lys-Pro-Arg, in water and in 1 *M* NaCl solution. In this case, the experimental NMR results less conclusively favor a type VI turn, and a separate series of simulations was used to study the *cis*- and *trans*-proline structures. Again, the *cis*-conformers tended to adopt compact conformations which are fairly rigid by typical peptide standards. The lysine and arginine side chains both tended to form intramolecular hydrogen bonds to the peptide backbone in water. The presence of salt had relatively little effect on the backbone structure, but did alter the average side-chain orientations. The *trans*-tuftsin simulation is dominated by more extended conformations.

8. Turns in β-Hairpin Structures

Several simulations show initial hydrophobic collapse followed by formation of turn structure, coupled in concert with detailed packing interactions. An early MD study ***(74)*** of the temperature unfolding of the residue 85-102 β-hairpin from barnase reported all-atom simulations from the native structure at 300 K, 450 K, and 600 K. The isolated hairpin in aqueous solution was stable over the course of 300 psec at 300 K, with an average root mean square from native of about 0.8Å for backbone atoms. Comparison of fluctuations in torsional angles during the 300 K hairpin simulation with those from a similar simulation of the full protein showed similar flexibility for the backbone angles. However, rotation about side-chain bonds was much more restricted in the full protein than in the isolated hairpin. Three persistent hydrogen bonds during the simulation involve the amide hydrogens of residues which are the most protected amide hydrogens during NMR hydrogen-exchange experiments. Following equilibration at 450 K, the structure is characterized by loss of hydrogen bonds at the extremities of the hairpin, more frequent exchange of hydrogen bonds with water, volume expansion, and formation of a large hydrophobic side-chain cluster containing both native and nonnative contacts. This loosely packed, yet still compact state was stable for the entire simulation 300 psec at 450 K, and for 150 psec at 600 K. Overall, the simulation data correlates well with current ideas generated from an experimental study of barnase unfolding.

In a recent report, Sung ***(75)*** used Monte Carlo sampling with a simple all-atom potential containing dielectric screened electrostatics and a surface area-based hydrophobic effect term to fold a capped poly-Val 12-mer with glycine substituted at the two central positions. Once again, the folding proceeded through an initial generalized collapse to a compact U-shaped structure as mediated by the hydrophobic effect. This same potential folds poly-Ala based sequences, which lack the potential for large hydrophobic stabilization, into α-helical motifs.

Many theories seem to favor a folding path directed by hydrophobic stabilization in which the turn does not need to be favorable or particularly stabilizing, but merely accessible and not prohibited. Recent thermodynamic decomposition of the experimental data shows hairpin folding to be an endothermic, entropy-driven process with a large negative ΔC_p, characteristic of a pathway dominated by formation of a hydrophobic core instead of specific enthalpically stabilized conformational preferences ***(76)***. Simulations of small regions of proteins using LINUS ***(77)***, however, suggest that folding is hierarchical and that there are biases towards local structures such as β-hairpins and β-turns primarily established by steric effects and hydrogen bonds and enhanced by hydrophobic interactions ***(78)***.

It is also likely that the situation is qualitatively different in larger globular protein systems. For example, Scheinerman and Brooks ***(79)*** have studied the free energy of folding of the 56-residue B1 segment of *Streptococcal* protein G along a reaction coordinate characterized by the percentage of native contacts. This globular domain is topologically similar to the much-studied chymotrypsin inhibitor 2 (CI2), and contains two β-hairpins separated in the primary sequence by an α-helical region. While the helix generally forms before the sheet structure, these researchers found that the earliest structure formed consisted of the N-terminus of the helix and the turn of the second hairpin. These two regions pack against each other in the native structure, and seem to independently form nucleation sites for the folding process.

9. Turns in Helix-Turn-Helix Structures

Sung and Wu ***(80)*** have used simulation to model the deformation of a long helix into various distorted helix-turn-helix conformations. The transition between the two forms occurs on a much longer time scale than simple helix propagation. A small model peptide similar to the central helix-turn-helix portion of ROP was folded using Monte Carlo dynamics ***(81)***. In the latter study, four of six trajectories ended in stable native-like conformations. The folding showed relatively fast formation of two helical regions of unequal length, followed by dynamic behavior in accord with the Karplus and Weaver diffusion-collusion model. Only in this latter stage did the sharp reverse turn develop, initiated by the formation of hydrophobic contacts at the end of the shorter helix. Once the two helices come into direct contact, the turn region expands by annexing some glycine residues from the last helical turn of the longer helix. In the final stable structure, the two helices are of nearly equal length. Both of these simulation studies are consistent with flexibility in the turn residues during the early stages of the folding event, with stable turn formation only after the helices are fully formed. They are also consistent with the experimental data discussed earlier.

10. Conclusion

There is no simple answer to the question posed at the beginning of this chapter. Computational studies reveal that small peptides which reflect turn motifs sample many conformational states, including those that ultimately appear in the final native structure. They also suggest that mutations in turn residues will certainly bias the amount of time spent in a particular local structure, and therefore the kinetic folding rates. Experimental evidence clearly shows that changes in turns or turn structures can affect both thermodynamic stability and kinetic properties. In addition to the turn itself, longer-range interactions between residues surrounding the turn are clearly important. It is surprising how few systematic experimental kinetic measurements have been made of the consequences of mutations in turns. Mutations that either include or replace glycine or proline residues may be particularly useful in this regard, but substitution between hydrophobic and hydrophilic residues or between those with different size side chains may be equally useful. One important result of such studies may be the possibility of creating long-lived intermediates. We propose turn scanning as a general method to quantitatively examine the importance of turns in the mechanism of protein folding with the understanding that such studies also require the concomitant use of biophysical techniques including theoretical simulation.

Acknowledgments

This work was supported by NIH Grant DK13332 (to Carl Frieden) and a Jane Coffin Childs Fellowship (to Enoch Huang).

References

1. Anfinsen, C. B., Haber, M., Sela, M., and White, Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **47,** 1309–1314.
2. Richardson, J. S. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34,** 167–339.
3. Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) Turns in peptides and proteins. *Adv. Protein Chem.* **37,** 1–109.
4. Fasman, G. D. (1989) Protein conformational prediction. *Trends Biochem. Sci.* **14,** 295–299.
5. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* **27,** 7167–7175.
6. Finkelstein, A. V. (1991) Rate of beta-structure formation in polypeptides. *Proteins* **9,** 23–27.
7. Blanco, F. J., Rivas, G., and Serrano, L. (1994) A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat. Struct. Biol.* **1,** 584–590.
8. Yang, A. S., Hitz, B., and Honig, B. (1996) Free energy determinants of secondary structure formation. 3. Beta-turns and their role In protein folding. *J. Mol. Biol.* **259,** 873–882.
9. Schonbrunner, N., Pappenberger, G., Scharf, M., Engels, J., and Kiefhaber, T. (1997) Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for β-sheet formation. *Biochemistry* **36,** 9057–9065.
10. Baldwin, R. L. and Rose, G. D. (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24,** 77–83.

11. Kim, K., Ramanathan, R., and Frieden, C. (1997) Intestinal fatty acid binding protein: a specific residue in one turn appears to stabilize the native structure and be responsible for slow refolding. *Protein Sci.* **6,** 364–372.
12. Zhou, H. X., Hoess, R. H., and DeGrado, W. F. (1996) In vitro evolution of thermodynamically stable turns. *Nat. Struct. Biol.* **3,** 446–451.
13. de Alba, E., Jimenez, M. A., and Rico, M. (1997) Turn residue sequence determines β-hairpin conformation in designed peptides. *J. Am. Chem. Soc.* **119,** 175–183.
14. Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390,** 196–199.
15. Munoz, V., Henry, E. R., Hofrichter, J., and Eaton, W. A. (1998) A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl. Acad. Sci. USA* **95,** 5872–5879.
16. Blanco, F. J. and Serrano, L. (1995) Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *Eur. J. Biochem.* **230,** 634–649.
17. Ramirez-Alvarado, M., Blanco, F. J., and Serrano, L. (1996) De novo design and structural analysis of a model beta-hairpin peptide system. *Nat. Struct. Biol.* **3,** 604–612.
18. Ramirez-Alvarado, M., Blanco, F. J., Niemann, H., and Serrano, L. (1997) Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J. Mol. Biol.* **273,** 898–912.
19. Blanco, F., Ramirez-Alvarado, M., and Serrano, L. (1998) Formation and stability of beta-hairpin structures in polypeptides. *Curr. Opin. Struct. Biol.* **8,** 107–111.
20. Schaefer, M., Bartels, C., and Karplus, M. (1998) Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **284,** 835–848.
21. Kim, K. and Frieden, C. (1998) Turn scanning by site-directed mutagenesis: Application to the protein folding problem using the intestinal fatty acid binding protein. *Protein Sci.* **7,** 1821–1828.
22. Ybe, J. A. and Hecht, M. H. (1996) Sequence replacements in the central beta-turn of plastocyanin. *Protein Sci.* **5,** 814–824.
23. Gu, H., Kim, D., and Baker, D. (1997) Contrasting roles for symmetrically disposed beta-turns in the folding of a small protein. *J. Mol. Biol.* **274,** 588–596.

24. Brunet, A. P., Huang, E. S., Huffine, M. E., Loeb, J. E., Weltman, R. J., and Hecht, M. H. (1993) The role of turns in the structure of an alpha-helical protein. *Nature* **364,** 355–358.
25. Predki, P. F., Agrawal, V., Brunger, A. T., and Regan, L. (1996) Amino-acid substitutions in a surface turn modulate protein stability. *Nat. Struct. Biol.* **3,** 54–58.
26. Castagnoli, L., Vetriani, C., and Cesareni, G. (1994) Linking an easily detectable phenotype to the folding of a common structural motif. Selection of rare turn mutations that prevent the folding of Rop. *J. Mol. Biol.* **237,** 378–387.
27. Peters, K., Hinz, H. J., and Cesareni, G. (1997) Introduction of a proline residue into position 31 of the loop of the dimeric 4-alpha-helical protein Rop causes a drastic destabilization. *Biol. Chem.* **378,** 1141–1152.
28. MacBeath, G., Kast, P., and Hilvert, D. (1998) Exploring sequence constraints on an interhelical turn using in vivo selection for catalytic activity. *Protein Sci.* **7,** 325–335.
29. Nagi, A. D., Anderson, K. S., and Regan, L. (1999) Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *J. Mol. Biol.* **286,** 257–265.
30. Viguera, A. R., and Serrano, L. (1997) Loop length, intramolecular diffusion and protein folding. *Nat. Struct. Biol.* **4,** 939–946.
31. Ladurner, A. G. and Fersht, A. R. (1997) Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* **273,** 330–337.
32. Nagi, A. D. and Regan, L. (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold. Des.* **2,** 67–75.
33. Helms, L. R. and Wetzel, R. (1995) Destabilizing loop swaps in the CDRs of an immunoglobulin VL domain. *Protein Sci.* **4,** 2073–2081.
34. Jackson, S. E. and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30,** 10,428–10,435.
35. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224,** 771–782.
36. Baldwin, R. L. (1996) On-pathway versus off-pathway folding intermediates. *Fold. Des.* **1,** R1–R8.
37. Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999) Structural changes in the transition state of protein folding: alterna-

tive interpretations of curved chevron plots. *Biochemistry* **38,** 6499–6511.

38. Santoro, M. M. and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27,** 8063–8068.
39. Hoeltzli, S. D. and Frieden, C. (1995) Stopped-flow nmr spectroscopy—real-time unfolding studies of 6-^{19}F-tryptophan-labeled escherichia coli dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA* **92,** 9318–9322.
40. Hoeltzli, S. D. and Frieden, C. (1994) 19F NMR spectroscopy of [6- ^{19}F]tryptophan-labeled E. coli dihydrofolate reductase: equilibrium folding and ligand binding studies. *Biochemistry* **33,** 5502–5509.
41. Hoeltzli, S. D. and Frieden, C. (1998) Refolding of [6-F-19] tryptophan-labeled Escherichia coli dihydrofolate reductase in the presence of ligand—a stopped-flow NMR spectroscopy study. *Biochemistry* **37,** 387–398.
42. Callender, R. H., Dyer, R. B., Gilmanshin, R., and Woodruff, W. H. (1998) Fast events in protein folding: the time evolution of the primary process. *Ann. Rev. Phys. Chem.* **49,** 173–202.
43. Plaxco, K. W. and Dobson, C. M. (1996) Time-resolved biophysical methods in the study of protein folding. *Curr. Opin. Struct. Biol.* **6,** 630–636.
44. Plaxco, K. W. and Dobson, C. M. (1998) Monitoring protein folding using time-resolved biophysical techniques, in *Protein Dynamics, Function and Design* (Jardetzky, O. and Lefevre, J., eds.), Plenum Press, New York, NY, pp. 163–172.
45. Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S., and Rousseau, D. L. (1997) Folding of cytochrome c initiated by submillisecond mixing. *Nat. Struct. Biol.* **4,** 44–50.
46. Shastry, M. C., Luck, S. D., and Roder, H. (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. *Biophys. J.* **74,** 2714–2721.
47. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. USA* **94,** 1779–1784.
48. Huang, G. S. and Oas, T. G. (1995) Submillisecond folding of monomeric lambda repressor. *Proc. Natl. Acad. Sci. USA* **92,** 6878–6882.

49. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) Microsecond protein folding through a compact transition state. *J. Mol. Biol.* **263,** 311–322.
50. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by ala-gly substitutions. *Nat. Struct. Biol.* **4,** 305–310.
51. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry* **37,** 5337–5343.
52. Arrington, C. B. and Robertson, A. D. (1997) Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry* **36,** 8686–8691.
53. Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1997) Fast events in protein folding—relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl. Acad. Sci. USA* **94,** 3709–3713.
54. Ballew, R. M., Sabelko, J., and Gruebele, M. (1996) Direct observation of fast protein folding—the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. USA* **93,** 5759–5764.
55. Nolting, B., Golbik, R., and Fersht, A. R. (1995) Submillisecond events in protein folding. *Proc. Nat.l Acad. Sci. USA* **92,** 10,668–10,672.
56. Nolting, B., Golbik, R., Neira, J. L., Solergonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. USA* **94,** 826–830.
57. Hutchinson, E. G. and Thornton, J. M. (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3,** 2207–2216.
58. Brooks, C. L. and Case, D. A. (1993) Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* **93,** 2487–2502.
59. Hermans, J. (1993) Molecular dynamics simulations of helix and turn propensities in model peptides. *Curr. Opin. Struct. Biol.* **3,** 270–276.
60. Tobias, D. J., Mertz, J. E., and Brooks, C. L. (1991) Nanosecond time scale folding dynamics of a pentapeptide in water. *Biochemistry* **30,** 6054–6058.
61. Hermans, J. and Anderson, A. G. (1990) Microfolding: use of molecular dynamics simulations to assess conformational stability of folded peptides and proteins in relation to protein engineering, in *Crystallographic and Modeling Methods in Molecular Design* (Bugg, C. and Ealick, S., eds.), Springer-Verlag, New York, NY, pp. 95–113.

62. Möhle, K., Gußmann, M., and Hofmann, H.-J. (1997) Structural and energetic relations between β turns. *J. Comput. Chem.* **18,** 1415–1430

63. Ösapay, K., Young, W. S., Bashford, D., Brooks, C. L., and Case, D. A. (1996) Dielectric continuum models for hydration effects on peptide conformational transitions. *J. Phys. Chem.* **100,** 2698–2705.

64. Yang, A. S., Hitz, B., and Honig, B. (1996) Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding. *J. Mol. Biol.* **259,** 873–882.

65. Lazaridis, T., Tobias, D. J., Brooks, C. L., and Paulaitis, M. E. (1991) Reaction paths and free-energy profiles for conformational transitions—an internal coordinate approach. *J. Chem. Phys.* **95,** 7612–7625.

66. Czerminski, R. and Elber, R. (1989) Reaction path study of conformational transitions and helix formation in a tetrapeptide. *Proc. Natl. Acad. Sci. USA* **86,** 6963–6967.

66a. Hart, R. K., Papp, R. V., and Ponder, J. W. (2000) Exploring similarities between potential smoothing and simulated annealing. *J. Comp. Chem.* **7,** 531–552.

67. Gunasekaran, K., Gomathi, L., Ramakrishnan, C., Chandrasekhar, J., and Balaram, P. (1998) Conformational interconversions in peptide beta-turns: analysis of turns in proteins and computational estimates of barriers. *J. Mol. Biol.* **284,** 1505–1516.

68. Yao, J., Dyson, H. J., and Wright, P. E. (1994) Three-dimensional structure of a type VI turn in a linear peptide in water solution. Evidence for stacking of aromatic rings as a major stabilizing factor. *J. Mol. Biol.* **243,** 754–766.

69. Mohanty, D., Elber, R., Thirumalai, D., Beglov, D., and Roux, B. (1997) Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water. *J. Mol. Biol.* **272,** 423–442.

70. Demchuk, E., Bashford, D., and Case, D. A. (1997) Dynamics of a type VI reverse turn in a linear peptide in aqueous solution. *Fold. Des.* **2,** 35–46.

71. Demchuk, E., Bashford, D., Gippert, G. P., and Case, D. A. (1997) Thermodynamics of a reverse turn motif. Solvent effects and side-chain packing. *J. Mol. Biol.* **270,** 305–317.

72. Jorgensen, W. L. and Gao, J. (1988) Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution. *J. Am. Chem. Soc.* **110,** 4212–4216.

73. Valdeavella, C. V., Blatt, H. D., and Pettitt, B. M. (1995) Simulations of conformers of tuftsin and a cyclic tuftsin analog. *Int. J. Pept. Protein Res.* **46,** 372–380.
74. Pugliese, L., Prevost, M., and Wodak, S. J. (1995) Unfolding simulations of the 85-102 β-hairpin of barnase. *J. Mol. Biol.* **251,** 432–447.
75. Sung, S. S. (1999) Monte Carlo simulations of beta-hairpin folding at constant temperature. *Biophys. J.* **76,** 164–175.
76. Maynard, A. J., Sharman, G. J., and Searle, M. S. (1998) Origin of β-hairpin stability in solution: structural and thermodynamic analysis of the folding of a model peptide supports hydrophobic stabilization in water. *J. Am. Chem. Soc.* **120,** 1996–2007.
77. Srinivasan, R. and Rose, G. D. (1995) LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins* **22,** 81–99.
78. Baldwin, R. L. and Rose, G. D. (1999) Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24,** 26–33.
79. Sheinerman, F. B. and Brooks, C. L., 3rd (1998) Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* **278,** 439–456.
80. Sung, S.-S. and Wu, X.-W. (1996) Molecular dynamics simulations of synthetic peptide folding. *Proteins* **25,** 202–214.
81. Hoffmann, D. and Knapp, E. W. (1997) Folding pathways of a helix-turn-helix model protein. *J. Phys. Chem. B* **101,** 6734–6740.

8

Laser Temperature-Jump Methods for Studying Folding Dynamics

James Hofrichter

1. Introduction

Until just a few years ago, kinetic studies of protein folding focused primarily on processes that occur on time-scales of milliseconds or longer—a limitation set by stopped-flow instrumentation. A number of factors motivated efforts to develop new experimental methods that could be used to characterize the kinetics of more rapid processes. First, a large body of theoretical work suggested that folding on the stopped-flow time-scale usually resulted from the escape of misfolded or partially folded structures from traps in the energy landscape. Folding by direct routes from the unfolded to the native state was suggested to occur much more rapidly ***(1–6)***. Second, the observation of significant unresolved amplitude in the fluorescence and circular dichroism signals in stopped-flow experiments suggested that significant structural organization could take place in the dead time of these experiments ***(7–12)***. Rapid formation of both native-like secondary ***(7–11)*** and tertiary ***(8–10,12)*** structure was observed within the stopped-flow dead time. Data on the dynamics of secondary structure formation for synthetic α-helical peptides ***(13,14)*** and on the rate of chain dif-

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

fusion under denaturing conditions ***(15–18)***, which suggested that helices and loops could form within microseconds, provided additional evidence that many interesting folding events occur on the submillisecond time scale.

What are some of the questions that may be addressed by "fast" folding experiments? To answer this question, it is helpful to consider an experiment in which the unfolding or refolding of a protein is initiated by a sudden change in the concentration of a denaturant (**Fig. 1**). For a protein exhibiting two-state thermodynamic and kinetic behavior, two processes are expected when the denaturant is diluted. The first is collapse of the unfolded polypeptide to more compact denatured structures under the new solvent conditions that favor folding (U′ → U), and the second is crossing the effective free-energy barrier from the new denatured state to the native state (U → N). When the concentration of denaturant is increased, structured regions which are less stable in the native state may unfold rapidly, (N → N′) followed by slower global unfolding (N′ → U′). While recent experiments using fast mixing methods may provide some information on the rate of collapse ***(19,20)***, many questions still remain. What are the dynamics of the denatured state? What limits the rate of collapse—the rates at which secondary structural elements form, or diffusion of the polypeptide backbone to form less local side-chain interactions? Does the chain initially collapse to a random distribution of topologies (a "random globule"), as suggested by lattice simulations ***(3)***, or to a set of compact structures with topologies similar to that of the native structure but without specific side-chain interactions (a "molten globule")? What is the height of the free-energy barrier separating native and denatured states, and how do we determine it from experimental rate constants? Is it possible to find systems that fold without a significant barrier?

To address these questions, we and others have developed a variety of new methods to more rapidly initiate folding and unfolding reactions. In this chapter, we focus on the development and application of the laser temperature-jump technique and describe how it has contributed to understanding the dynamics of protein folding.

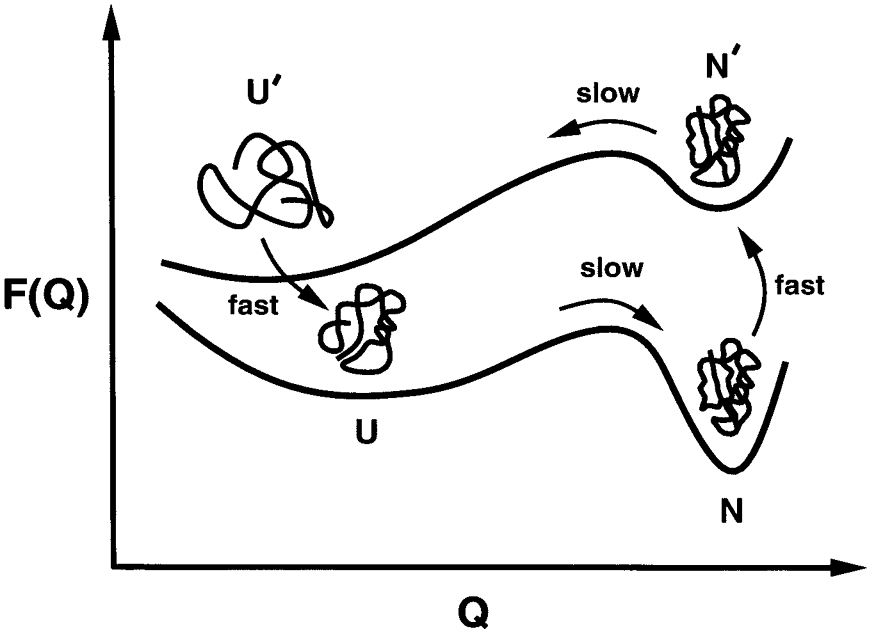


Fig. 1. Free-energy profiles for folding and unfolding of a two-state protein by a change in the concentration of a chemical denaturant.

In our laboratory at the National Institutes of Health (NIH), we have used temperature-jump (T-jump) methods to begin a “bottom-up” approach to the folding problem. The idea is to experimentally characterize the dynamics of the simplest elements of protein structure—α-helices, β-hairpins, and loops—and to develop statistical mechanical models to describe these dynamics. Using this approach, we have investigated the dynamics of three basic elements of protein structure in isolation—loops ***(16–18)***, α-helices ***(21–23)***, and β-hairpins ***(24)***. The experimental results have provided the first glimpse of the rates of these three fundamental dynamical processes in protein folding, and the theoretical modeling has provided a means for exploring how these rates are related to structure and sequence. Other groups have used this approach to study rapid processes which accompany the thermally induced folding ***(25–27)*** or unfolding of proteins ***(28–30)***.

We will first discuss the experiment in general terms, beginning with a brief historical perspective on the development of the laser temperature-jump technique, followed by a review of the basic physics of the experiment, including some discussion of some of the problems that may be encountered in carrying out such experiments. Some instruments that have been used in studying protein folding are described, and the experimental findings are briefly summarized.

2. The Laser Temperature-Jump Experiment

2.1. Historical Overview

Beginning with the landmark work of Eigen and colleagues ***(31)***, it has been recognized that temperature changes provide a general and practically useful means of disturbing a system in order to perform rapid kinetic studies. It is important to realize that temperature is one of a relatively large number of perturbants—including pressure, electric fields, and changes in solution conditions—which can be applied to this purpose. Since temperature jumps are almost always positive, in the study of the kinetics of protein folding they can be used both to study the rapid unfolding to thermally denatured states ***(25–27)*** and the rapid folding of proteins from cold-denatured states ***(28–30)***.

For more than a decade, temperature-jump studies provided one of the most useful means for extending dynamics studies to timescales shorter than milliseconds. The early work was carried out using electical current pulses, which were generated by discharging a capacitor by triggering a spark gap ***(31)*** to obtain rapid heating. This approach required the use of solutions of high conductivity and thus high ionic strength. After the advent of commercial pulsed lasers in the late 1960s, laser flash photolysis soon became the most commonly used method for carrying out fast kinetics studies. Efforts were also made to use pulsed lasers to develop practical laser temperature-jump instruments ***(32–36)***. In a laser temperature-jump experiment, a pulse of laser energy is absorbed by a dye or by the solvent, and undergoes radiationless decay to produce heat. It was clear from the beginning that by using a laser pulse, the temperature change could be produced within a few nanoseconds using a Q-switched laser ***(32)***. The shorter pulses obtained from mode-locked lasers would have to be pulse-selected and amplified to produce significant temperature jumps, but these would allow resolution on the time-scale of a few picoseconds or less ***(32)***. Initial efforts utilized ruby and neodymium glass lasers—the only commercial lasers to generate enough pulse energy to produce temperature jumps of >1° ***(32)***. These lasers produce radiation at 694 nm and 1060 nm, respectively, where the absorption of water is negligible and very weak, so many of the early studies utilized absorbing dyes to produce the temperature jump ***(32,33)***.

There are a number of advantages to using a near-infrared laser pulse to directly heat the aqueous solvent. First, it eliminates both the absorption of the dye and its photophysics as potential problems in designing a practical temperature-jump experiment. Second, because the heated vol is determined by the laser spot size and the solvent A, it can be varied from as large as a few mm^3 to sizes as small as a few cubic microns, a fraction of the vol of a living cell. Third, the use of direct absorption by the solvent permits complete flexibility in determining solution conditions without the necessity of working with the high salt concentrations necessary for electrical temperature-jumps. Finally, the use of IR radiation minimizes interference between the pump pulse and optical detection of the sample response. A wide variety of probe signals, including visible and IR absorption, fluorescence, and Raman scattering, can therefore be used to probe the dynamic response of the heated sample.

The generation of large temperature jumps by directly exciting water requires pump pulses with energies of several mJ at wavelengths between 1300 and 2100 nm. The first successful experiments utilized the stimulated Raman effect in liquid nitrogen to shift Q-switched pulses from a Nd:YAG laser from the 1.06 μm fundamental to 1.41 μm ***(34,35)*** where the A of water is approx 10 cm^{-1}. Difficulties encountered in using liquid N_2 (unstable pulses caused by boiling at room temperature, self focusing, and the need for a very large dewar) prompted a search for more convenient Raman active media. Stimulated Raman scattering from H_2 gas, which produces a frequency shift from 1.06 µm to 1.89 µm, was successfully used in one early study ***(36)***. The cross-section of water at 1.89 µm is approx 30 cm^{-1}, so the path lengths of the sample are limited to less than about 0.2 mm. It was more than a decade after these pioneering studies, however, before Turner and colleagues ***(37,38)*** first utilized Raman shifting in methane for this application. Methane shifts the Nd:YAG fundamental at 1064 nm by 2917 to 1540 nm, where the water A, *k*, is about 12 cm^{-1}.

There are a number of other methods for generating suitable IR pump pulses. Holzwarth et al. ***(39)*** first used an iodine laser with an output at 1.315 µm to heat water directly, a technique which is still in use today. More recently, holmium and erbium lasers—operating at 2.065 µm and 1.54 µm, respectively—have become commercially available. The design and construction of a non-Q-switched holmium laser, used to produce temperature jumps in muscle fibers ***(40)*** has been described ***(41)***, and commercial Q-switched holmium lasers are available. Tunable (IR) radiation is also achievable by difference mixing as described in **Subheading 2.2.** ***(42)***, and by using Nd:YAG or Nd:YLF lasers to pump optical parametric oscillators. One advantage of tunable IR is that the pump wavelength can be selected to optimize the temperature jump for different sample path lengths, which can be chosen to optimize the parameters of the probe experiment.

2.2 Physical Background

The temperature change produced by the laser pulse, δT, in the absence of thermal diffusion, is given by:

$$\delta \mathrm{T}(r,z,t) = k/\rho \mathrm{c_v} \int_0^t I(r,z,t')\,\mathrm{d}t' \qquad (1)$$

where $I(r)$ describes the intensity profile of the pulse and $I(r,z,t)$ is the instantaneous laser fluence in W/cm^2 at time *t* and axial position *z*, *k* is the absorption cross-section per U length, and ρ is the solution density, and c_v is its heat capacity. In an absorbing medium, the

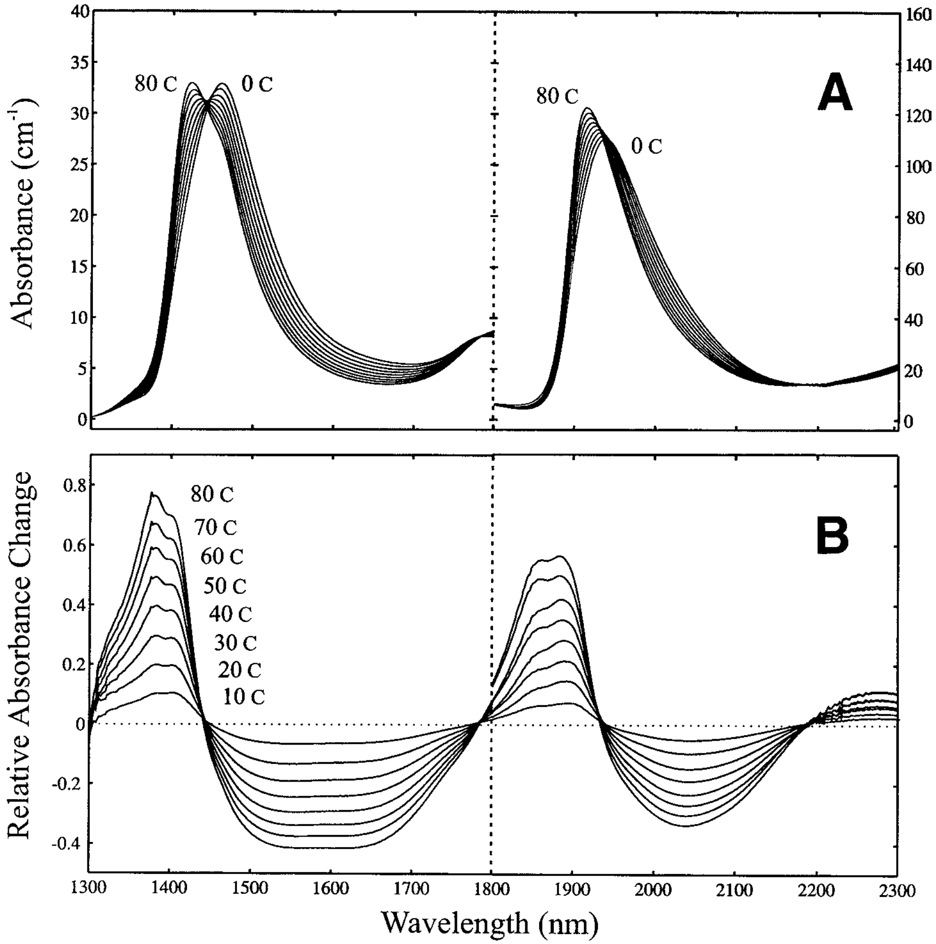


Fig. 2. Near IR absorption of water. **(A)** The A, *k* (**Eq. 2**), of water at temperatures from 0°–80°C, measured using an Aviv 14 spectrophotometer. **(B)** Fractional change in A (relative to the A at 273 K) at temperatures from 10°–80°C.

dependence of intensity on the spatial coordinate, *z*, is given by the Lambert Law:

$$\frac{\partial I(r,z)}{\partial_z} = -kI \; ; \qquad I(r,z) = I_0(r)\exp(-kz) \qquad (2)$$

The A of water in the near IR is shown in **Fig. 2** and the A of D_2O is shown in **Fig. 3**. For a pulse at 1540 nm, where the A, *k*, is about

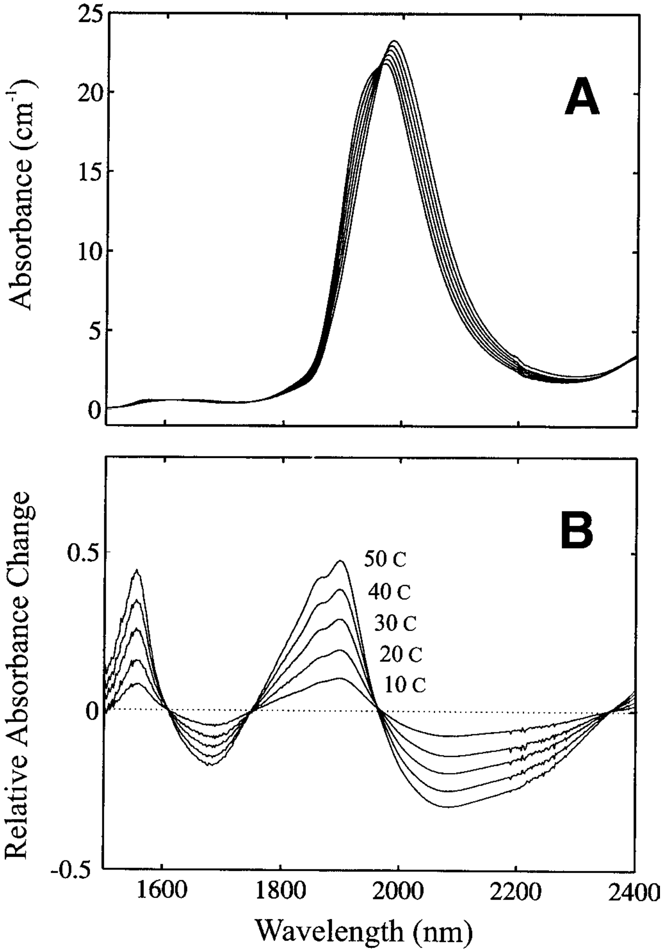


Fig. 3. Near IR absorption of D_2O. **(A)** The A, k (**Eq. 2**), of D_2O at temperatures from 0°–50°C, measured using an Aviv 14 spectrophotometer. **(B)** Fractional change in A (relative to the A at 273 K) at temperatures from 10°–50°C.

12 cm^{-1} at 293 K, the temperature rise at the surface of the cuvet is therefore about 1 K for an integrated pulse energy of 1 mJ/mm^2. Note that the spectra of both H_2O and D_2O are significantly temperature-dependent. For H_2O, the A is nearly temperature-independent at wavelengths of 1.44 μm, 1.94 μm, and 2.18 μm, while for D_2O the A is nearly temperature-independent only at

1.96 μm. At other wavelengths the A varies by as much as 1% per degree. At 1.54 μm, the A of H_2O decreases by about 0.5% per degree, from 14.5 cm^{-1} at 273 K to only about 8.5 cm^{-1} at 353 K, so temperature jumps at 353 K are about 40% smaller than those produced at 273 K. At 2.0 μm, where a number of IR experiments have been carried out in D_2O, the temperature dependence is similar.

The duration of the temperature pulse is limited by thermal diffusion. In a cell with a path-length which is long compared to the beam diameter, the decay time can be roughly estimated by considering the problem of an infinite cylinder which is uniformly heated relative to the surrounding medium at time t = 0. The temperature at the origin is given by:

$$\delta T(t) = \delta T_{\max}(1 - \exp(-r^2 / 4\kappa t) \quad (3)$$

where r is the radius of the cylinder and κ is the thermal diffusivity in cm^2/s ***(43)***. The value of κ for water at 293 K is $1.43(10^{-3})$ cm^2/s. The usable duration of the temperature pulse can therefore be crudely estimated using $t = r^2/12\kappa$ to give the time at which the temperature has decayed by ~5%. For the 1-mm^2 spot which we used as an example above, this decay time is about 0.1 s. For cells which are thin compared to the beam diameter, the temperature can be roughly estimated by the expression for a uniformly heated slab in a homogeneous medium:

$$\delta T(z,t) = \delta\, T_{\max}\left(erf\,\frac{a-z}{2\sqrt{\kappa t}} + erf\,\frac{a+z}{2\sqrt{\kappa t}} \right) \quad (4)$$

where 2a is the cell path-length and z is the distance from the center of the cell ***(43)***. In this case, the duration of the pulse is determined by the path-length, a, and is nearly independent of the illuminated area. Since almost all of the heat flows through the windows, the value of κ must be taken as that of the window material, not of the solution, in using **Eq. 4** to estimate the pulse duration. In real experimental geometries, the accurate calculation of the temporal decay of the pulse is significantly more complex, since the sample is not uniformly heated or homogeneous. In most cell geometries, significant conduction occurs through the windows of the cuvet, which usually has a thermal diffusivity significantly larger than that of water, and an accurate solution of the diffusion problem requires that it be solved by finite element integration.

Equations 1–4 contain the basic information required for engineering a temperature-jump experiment. The size of the temperature jump depends inversely on the illuminated area, and the the duration of the pulse is determined by the cell and beam geometry. Given the excitation wavelength and the A from **Figs. 2** or **3**, one can determine the optimal path-length. If the temperature jump is more than a few degrees, it is important to minimize the gradient in temperature across the area that will be probed. For single-pass excitation, if the temperature jump is to be constant to within 10%, the A of the sample must be less than approx 0.04 optical density (OD). To use the excitation energy more efficiently, it is necessary to utilize a multiple-pass geometry. One simple arrangement is to reflect the attenuated pulse back through the sample. In this geometry, the A can be increased to about 0.2 OD. An even better approach is to split the excitation beam using a 50/50 beam-splitter and excite from both sides ***(24,32,44)***, in which case the A can be as large as 0.4 OD. The increased ODs make use of a larger fraction of the incident pulse energy, so despite the decrease in incident energy which results from splitting the beam, the size of the largest homogeneous temperature jump produced using either of the latter two geometries is almost identical.

The diameter of the pulse is determined by two factors. The first is the required duration of the temperature jump, and the second is the probe optics. If the temperature must be constant to within 5% of the jump amplitude for time *t*, then in the thick-cell limit, the size of the spot can be estimated using **Eq. 3**: a 1 mm^2 spot can be used to measure processes as long as about 100 ms, an 0.1-mm^2 spot for processes as long as about 10 ms, and an 0.01 mm^2 spot for processes as long as about 1 ms. The required laser-pulse energy can then be calculated from **Eq. 1**. For the 1 mm^2 example, pulse energies of ~20 mJ are required to produce temperature jumps of approx 10 K. A more important consideration is the requirement that the experiment be probed over an area with a temperature that is nearly constant. For a gaussian pulse profile, the requirement that the temperature jump be constant to within 10% restricts the probe-beam diameter to approx 0.3 of the 1/e diameter of the pump beam. More important, however, is the consequence of thermal lensing, by even this small fraction of the total temperature gradient produced in the sample (*see* below). In practice, we have found that the use of a beam with approx 0.2 of the pump-beam diameter minimizes lensing artifacts for temperature jumps of 15 K or less in water ***(21–23)***.

2.3. Thermal Lensing and Photoacoustic Effects

The rapid expansion of the heated vol produced by the temperature jump has three consequences, which can interfere with the probe signals measured in laser-temperature jump experiments. These are thermal lensing, the generation of photoacoustic waves, and cavitation ***(32,35,44–46)***. Both thermal lensing and photoacoustic waves produce changes in the spatial distribution of probe-beam intensity as it passes through the heated sample, which can result in artifactual changes in the intensity of the measured probe signal. The rapid expansion of the heated vol after the temperature jump can produce cavitation, particularly in solvents of low vapor pressure. Because all of these effects are directly related to changes in the density of the solution, all three problems can be minimized by working under conditions where no change in sample vol is produced by the temperature jump. In aqueous solutions, this can only be accomplished by carrying out jumps at temperatures close to 277 K where the density is maximal. It is clear that this solution to the problem imposes unacceptable constraints on the temperature-jump technique if it is to be used as a general tool for studying protein folding. Therefore, it is important to understand these effects and delimit conditions that minimize their interference with the experiments.

For short laser pulses, the temperature change occurs at constant volume because the inertia of the surrounding fluid prevents the heated fluid from expanding. At temperatures above the density maximum (277 K), the temperature jump produces a relatively large increase in pressure in the heated volume. Prior to expansion, the pressure rise is given by:

$$\left(\frac{\partial P}{\partial T}\right)_V = \left(\frac{\partial V}{\partial T}\right)_P \Big/ \left(\frac{\partial V}{\partial P}\right)_T = \alpha/\beta \tag{5}$$

where α is the coefficient of isobaric expansion and β is the isothermal compressibility of the medium. For water at 293 K $\alpha = 2.0\ (10^{-4})\ \mathrm{K}^{-1}$ and $\beta = 45.3\ (10^{-6})\ \mathrm{bar}^{-1}$, so the pressure rise is 4.4 bar/K. The pressure pulse then propagates outward from the heated vol as a shock wave. In the process, the pressure in the liquid surrounding the heated vol increases and the surrounding fluid is accelerated outward. The duration of these events is determined by the speed of sound, which is the propagation velocity of the pressure pulse. This photoacoustic wave can be reflected back into the sample vol by changes in the speed of sound which occur at the walls of the cuvet. If significant pressure pulses are reflected back into the probed vol, they can produce large changes in the refractive index and thus the measured intensity of the probed signal. In most cell geometries, these effects occur on the µs time scale ***(32)***. The outward momentum of the fluid layer immediately surrounding the heated vol also produces a negative (tensile) pressure pulse on the sample vol. This negative pressure, which can last ~100 ns ***(45)***, is the most probable source of cavitation. Studies have examined the formation of bubbles in water to which absorbing dyes have been added, and cavitation has been found to occur when the negative pressure in the heated vol reached 5–10 bar ***(45,46)***. Bubbles with maximum diameters of 25–100 µm and lifetimes of a few ms were produced.

Expansion of the heated vol takes place much more rapidly than thermal diffusion. As the sample expands, the density of the illuminated region of the cell decreases, producing refractive-index gradients that focus or defocus the probe beam. The effects of focusing on the measured signal depend on the instrument geometry and the signal being measured. Because the thermal lens alters the beam profile at some distance from the sample more than within the sample cell itself, the contributions of lensing to the measured signal produce more significant problems for absorption measurements than for fluorescence or other emission experiments. In extreme cases with small detectors, the thermal lens can focus a significant fraction of the transmitted intensity off the detector, producing large apparent increases in OD. Even when the area of the detector is larger, thermal lensing redistributes the transmitted light on the surface of the detector, and thus may still produce small changes in the apparent A unless the sensitivity of the detector is extremely homogeneous.

3. Laser Temperature-Jump Instruments

Several temperature-jump instruments have been developed for the study of protein folding. These instruments differ in many ways, including the means for producing the temperature jump and in the signal detected. In **Subheading 3**, we describe three instruments: the instrument in our lab at the NIH, the instrument built by Brian Dyer at Los Alamos National Laboratories, and the instrument developed in Martin Grubele's laboratory at the University of Illinois at Urbana-Champaign.

3.1 The NIH Instrument

Our instrument was designed to use either absorption or fluorescence in the visible or near-UV as a probe signal. A 1.54-µm IR pulse, produced by Raman-shifting the fundamental of a Q-switched Nd:YAG laser, is used to rapidly heat the sample, and a continuous wave (cw) laser beam is used to probe absorption or to excite the fluorescence of the sample and to monitor the relaxation kinetics. In the peptide experiments discussed here, a 264-nm UV probe beam is used to excite the fluorescence of tryptophan. A schematic of the instrument is shown in **Fig. 4**. To produce the temperature-jump pulse, the fundamental (1064 nm) of a Nd:YAG laser (Continuum Surelite I), operating at 1.67 Hz, is focused with a 0.75-m lens into a 1-m Raman cell (Princeton Optics, Inc.). The Raman cell contains

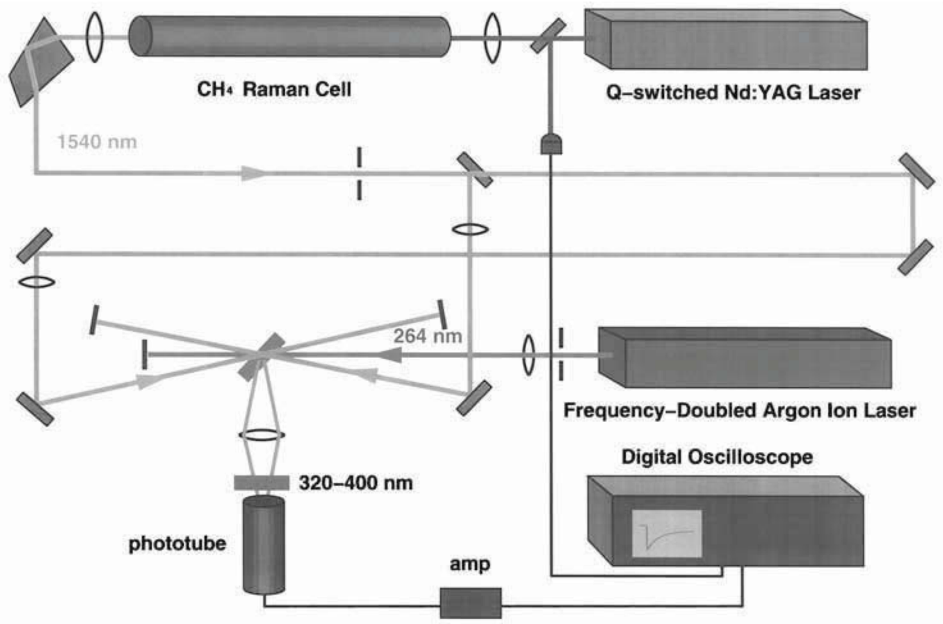


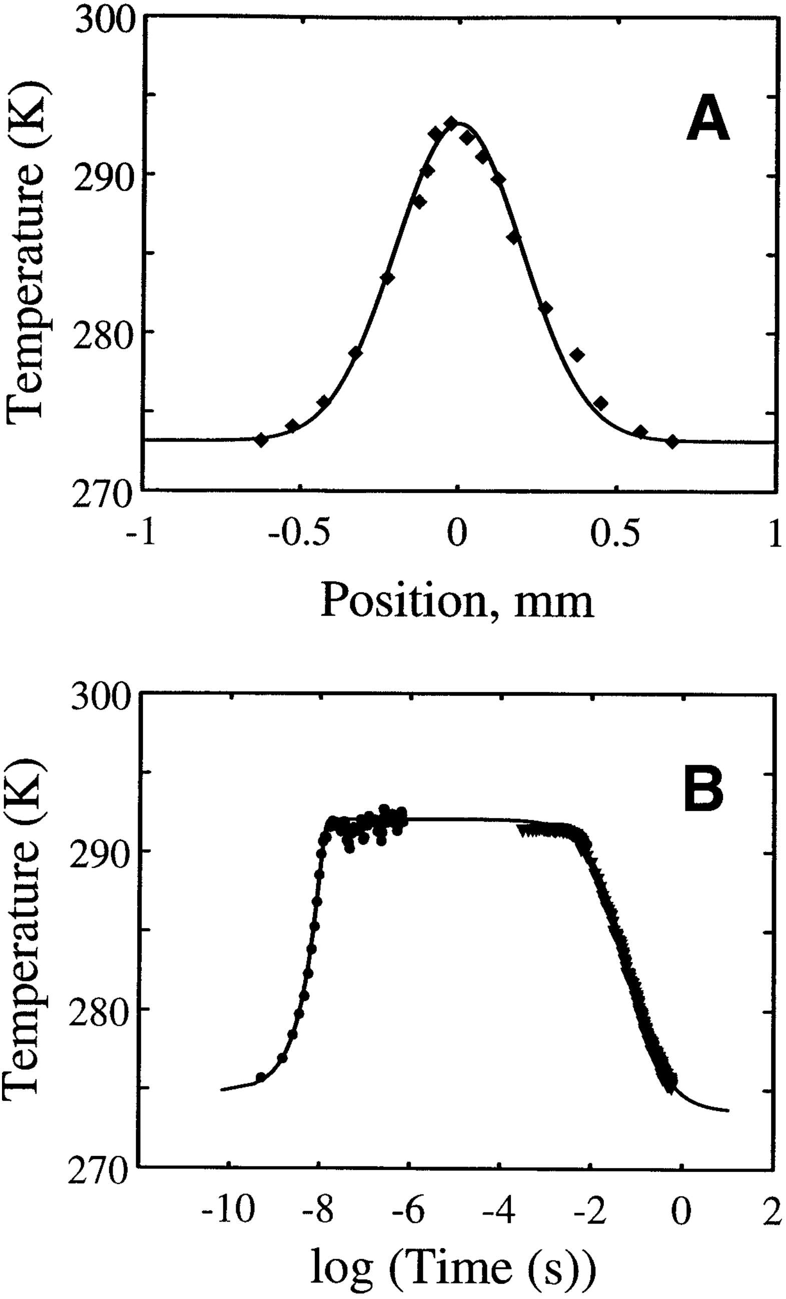
Fig. 4. Schematic of the temperature-jump instrument. An ~5 ns (FWHM) temperature-jump pulse is generated using the stimulated Raman effect to shift the Nd:YAG fundamental (1064 nm) to 1.54 µm in a mixture of CH_4 and He gas. 10 mJ of 1.54 µm light is split into two beams which are focused onto the 500 µm path-length sample cell to give a spot size of 0.5 mm. A cw UV beam at 264 nm from an intracavity-frequency-doubled argon ion laser is focused to a spot size of 60–70 µm to excite the tryptophan chromophore. The fluorescence signal is collected at 90° from the excitation beam, filtered using a bandpass filter, and measured using an end-window photomultiplier tube.

600 psi of CH_4 and 500 psi of He, and has a conversion efficiency of up to 20% for the first Stokes line (1.54 µm). The 1.54-µm pulse, which has a temporal width of ~5 ns, is separated from the fundamental and anti-Stokes lines with a Pellin-Broca prism, and then split with a 50:50 beam-splitter. The two resulting pulses are routed toward the sample from opposite sides, with the length of the optical path for one approx 2 m longer than the other. The pulses from each arm are focused onto the same spot of the sample, using an

0.75-m lens to produce a spot diameter of 0.5 mm. As a result of the difference in optical path, one pulse is delayed by ~7 ns relative to the other, increasing the rise time of the resulting temperature jump, and thus reducing the probability of photoacoustic interference in the probe signal. We have not encountered serious problems from photoacoustic reflections, possibly because the flat-cell geometry helps to dissipate the pressure pulse before it encounters the walls of the cuvet. Using a 500-µm path-length cuvet, approx 40% of the IR beam is absorbed by the sample, producing a temperature rise of approx 2 K per mJ of IR pulse energy.

To probe tryptophan fluorescence, an ultraviolet (UV) beam is generated with an intracavity frequency doubled argon ion laser (FRED, Coherent Inc. Santa Clara, CA). This laser provides a series of lines between 229 nm and 264 nm. The excitation beam is focused onto the sample with a 10-cm lens to produce a spot size of 60–70 µm. For fluorescence experiments, a front-face illumination geometry is used in which the thin sample cell is oriented at 45° to the pump and probe beams. The fluorescence signal is collected at 90° from the excitation beam using a 3.9-cm focal length lens, filtered from residual reflected excitation and IR light using bandpass filters which transmit wavelengths from 320–400 nm, and detected with a photomultiplier tube. To minimize lensing signals, we use an end-on photomultiplier tube (Hamamatsu R6427) with a large active photocathode area with reasonably uniform cathode sensitivity. The signal is amplified by cascading two stages (5x gain per stage) of a fast preamp module (Stanford Research Systems, SR240) and processed by a digital oscilloscope (Tektronix, TDS 620). The waveforms from the oscilloscope are transferred to a personal computer for analysis. Typical experiments require averaging of 10^3–10^4 laser shots.

We have used a number of methods to measure the magnitude of the temperature jumps. For tryptophan experiments we measure the change in the fluorescence intensity of n-acetyl tryptophan amide (NATA), which decreases about 1% for each degree of temperature increase. The intensity change is compared with the results of a static temperature scan for the same sample. **Figure 5B** shows that


A
Temperature (K)
300
290
280
270
-1
-0.5
0
0.5
1
Position, mm
B
Temperature (K)
300
290
280
270
-10
-8
-6
-4
-2
0
2
log (Time (s))

the instrument response time for a temperature jump from 0°C to 20°C is ~5 ns, and that thermal diffusion from the reaction vol takes several milliseconds. When working in the absorption mode, we found the absorption change of tyrosine in unbuffered solution at pH 10.0 to be useful as a spectroscopic thermometer.

3.2. The Los Alamos Instrument

A somewhat different approach has been used by the group at Los Alamos National Labs to carry out time-resolved mid-IR experiments which probe the secondary structure of peptides and proteins. For experiments on the ns time-scale, the pump beam is generated by mixing the fundamental of an injection-seeded, Q-switched Nd:YAG laser at 1064 nm with the 683-nm output of a pulsed dye laser which is pumped with the second harmonic of the same laser. The dye LDS 698 is used to produce the dye pulse and a $LiNb0_3$ crystal in an IR WEX (Spectra-Physics, Mountain View, CA) is used to produce the difference frequency at 2.0 µm. This wavelength was chosen because the experiments are carried out in D_2O, which minimizes interference with the IR measurements of the amide I bands between 1600 and 1700 cm^{-1}. D_2O has an A of approx 20 cm^{-1} at this wavelength (**Fig. 3**), approximately twice that for water at 1.54 µm. The higher A produces larger temperature jumps (*see* **Eq. 1**); heating a 300-µm diameter spot with an integrated pulse

---

Fig. 5. *(opposite page)* Spatial and temporal characterization of the temperature jump. **(A)** Spatial profile of the temperature pulse. The fitted diameter at half-peak height is 0.49 mm. **(B)** Temporal profile of the temperaure jump. The rise in temperature is fitted as the convolution of a pulse with a gaussian temporal profile and a half-width of 12 ns. The decay of the pulse is fitted using the function $T(t) = T_0 + \Delta T\tau/(t + \tau)$ with $\Delta T = 18.4$ K and $\tau = 61$ ms. The temporal dynamic range of the temperature-jump instrument extends from ~ 10 ns, which is twice the apparent rise time to the final temperature, to ~ 5 ms, the point at which the temperature has decayed by 10% of the temperature change, ΔT.

energy of 2 mJ produces temperature jumps of 15–20 K. A cell with a path-length of 0.1 mm is used for the IR absorption measurements. Because the cell is thinner than the beam diameter, conduction from the heated vol occurs primarily through the CaF_2 windows, the thermal diffusivity of which is about 30 times greater than that of D_2O. Consequently, the duration of the pulse is only about 1 ms, and the temperature is constant to within 10% for approx 0.2 ms.

Infrared absorption is probed with a CW IR diode laser, which can be temperature-tuned from 1632 cm^{-1} to 1700 cm^{-1}. The intensity of the probe beam is detected by an HgCdTe detector plus a preamplifier, and digitized with an oscilloscope. The time resolution of the system is limited by the bandwidth of the detector-amplifier combination to about 20 ns. By averaging about 10^4 shots, instrument sensitivity of about $2(10^{-4})$ OD is achieved ***(42)***.

3.3 The UI-Urbana Instrument

An elegant laser-temperature-jump instrument with time-resolved fluorescence detection has been constructed by Martin Grubele and colleagues at the University of Illinois at Urbana-Champaign ***(28–30,47)***. The excitation pulse is produced by using a 750-mJ input pulse from a Q-switched Nd:YAG laser at 1064 nm to pump a 1.75-M Raman cell in a collimated geometry. The methane Raman cell is operated at 21–22 bar, producing ~200 mJ of 1.54 μm IR. As at the NIH, the output pulse is split after isolation from the pump pulse, and the two pulses are focused to a diameter of ~1 mm on the 0.5-mm path-length sample cell. The large beam diameter and cell thickness combine to produce a pulse with a decay time of about 0.4 s and a useful duration of about 100 ms. The size of the temperature jump and its duration are monitored by directly measuring the temperature-dependent A of water at 1.5 μm (*see* **Fig. 2**). A chopped diode laser is used as a source, and a Germanium photodiode coupled to a lock-in amplifier is used to measure the transmitted intensity. While the rise time of the temperature jump cannot be directly monitored using this approach, it provides an important control on variations in its magnitude.

To probe the fluorescence decay, a train of pulses from a mode-locked titanium:sapphire laser is used to measure a series of flourescence decay curves after a ns temperature jump. The emission of tryptophan, excited at 280 nm by the tripled output of the Ti-sapphire laser, is monitored by a fast photomultiplier-amplifier combination with a rise time of 0.6 ns. To obtain accurate fluorescence decay curves through the entire record, which can be as long as tens of milliseconds, the Ti:sapphire laser is operated at 62.5 MHz, producing a pulse every 16 ns. The eighth harmonic of this frequency is used to clock the digitizer at 500 MHz, thereby ensuring that the excitation pulse and the sampling points of the digitizer remain precisely phased through the entire record ***(47)***. If this approach is not used, aliasing would make it difficult to compare the fluorescence decay kinetics measured at different times after the temperature jump. (The extremely long record lengths currently available in digital oscilloscopes make this approach feasible with sampling rates as high as 4 GHz). The set of decays is measured for a single shot, so the repetition rate of the pump laser is not an issue. To analyze this data, the singular value decomposition of the entire set of fluorescence decay curves is calculated. The set of coefficients which describe the resulting amplitudes is then fitted as the linear combination of two decay curves: the decay curve prior to the temperature jump, and the equilibrium decay curve at the higher temperature. Because the lifetime of tryptophan is highly temperature-dependent, the bulk of the difference results from the change in temperature ***(28,29)***, but the effects of structural changes on the fluorescence decay are observed at times after the temperature change is complete.

4. Studies of Folding Kinetics

Because the application of laser temperature-jump techniques to protein folding is relatively recent, it is possible to briefly summarize the key results.

4.1. Peptides and Model Systems

One important application of the laser temperature-jump methods is to study the formation of specific secondary structural elements that are formed by peptides. Some of these systems, such as the C-terminal hairpin of protein G β1 studied by Muñoz et al. ***(24)***, are sequences that have been excised from protein sequences ***(48)***. Others, such as the alanine helices originally described by Baldwin and colleagues ***(49–51)***, are synthetic peptides designed to fold into specific secondary structures. Because of the unavailability of both experimental methodology and suitable model systems, the dynamics of secondary structure formation were largely unknown as little as five years ago.

Helix melting in short peptides was first studied by Williams et al. ***(27,42)*** using the Los Alamos instrument described here. Using the 21-residue peptide Suc-AAAAA-$(AAAR^+A)_3$-A-NH_2, (where A=alanine, R=Arginine and Suc=succinyl) and probing using IR A at wavelengths between 1632 cm^{-1} and 1664 cm^{-1}, they monitored the disappearance of the amide I A of α–helical residues and the appearance of the A of random coil residues. A relaxation time of 160 ± 60 ns was measured—a result which is reasonably consistent with earlier studies of large model polypeptides in water ***(52)***. The experiments were carried out in D_2O at 301 K, a temperature well below the midpoint of the melting transition—which is reported as 333 K in D_20 ***(42)***, considerably higher than the 306 K reported in water ***(53)***. This seminal work was followed by attempts to utilize absorption and fluorescence probes of helix-coil dynamics. Thompson et al. ***(21,22)*** used a 21-residue peptide MABA-AAAAA-$(AAAR^+A)_3$-A-NH_2, (where MABA = methyl-amino benzoic acid) originally studied by Lockhart and Kim ***(53,54)***. A significant increase in the fluorescence of MABA was observed in the folded state, and the fluorescence decreased with a rate of about 10^8 s^{-1} after the temperature jump. The relaxation rate, which was about a factor of 2 slower than the time resolution of the instrument, was nearly temperature-independent. The rapid relaxation was attributed to unzipping of the N-terminal end of the helix, and the fluorescence change was attributed to breaking of the helical hydrogen bond formed between MABA and residue A_4.

We recently used a better-characterized probe to study the helix-coil transition of the heteropeptide Ac-WAAAH$^+$(AAAR$^+$A)$_3$A-NH$_2$ ***(23)***. This peptide was designed with tryptophan in position 1 and histidine in position 5, so that the side chains interact when the backbone of residues 1–5 is α-helical. The histidine, when protonated, efficiently quenches tryptophan fluorescence, providing a probe for the presence of helical structure. The kinetics measured throughout the melting transition are well-described by a single exponential relaxation, with a rate of 3.3×10^6 s^{-1} at 301 K, the midpoint of the helix-coil transition. The rate increases with increasing temperature with an apparent activation energy of approx 8 kcal/mol, so the relaxation time varies from ~500 ns at 280 K to less than 100 ns above 320 K. The equilibrium and kinetic data were fitted using a statistical mechanical model developed by Muñoz et al. ***(55)***. The model successfully explains the single-exponential kinetics of the fluorescence change as resulting from the interaction between the tryptophan and the charged histidine in position 5. This interaction acts as a "staple" which tends to lock the first five residues into the helical conformation. As a result, the tryptophan rarely melts unless the entire helix melts, and approx 90% of its fluorescence change is predicted to occur through the melting of complete helical sequences. The measured temperature-dependence can be reproduced by introducing damping from solvent friction ***(56)***, and an activation barrier for the individual helix propagation steps into the rates for the helix propagation steps.

The first information on the dynamics of β structure formation was obtained by Muñoz et al. ***(24)*** using a similar approach. This study used a protein fragment consisting of residues 41 to 56 of Streptococcal protein G B1. The sequence contains a single tryptophan (W43), which provides an intrinsic probe of hairpin formation. The W43 fluorescence increases in the folded conformation because it forms a hydrophobic cluster that includes phenylalanine 52 (F52) and valine 54 (V54). The hairpin acts as two-state system (i.e., the peptide is either unfolded or has a conformation with the β-hairpin environment for W43), and unfolding occurs in a single exponential relaxation with a rate which is about 20 times slower than the helix-coil transition for an alanine-based peptide. The measured relaxation times increase from about 4 μs at 285 K to about 1 μs at 330 K. When a dansylated lysine was added to the C-terminus of the peptide, the fluorescence in the folded state decreased as a result of excitation energy transfer from the tryptophan to dansyl, but the observed relaxation rates were almost identical. Using a two-state analysis, the activation energy for the folding transition is *negative*.

Muñoz et al. were able to explain all of these findings using a simple statistical mechanical model ***(25,55)***. In the case of helical alanine peptides, the top of the barrier between the unfolded and folded peptide is reached after fixing three peptide bonds. Elongation is then thermodynamically favorable, because the entropy loss of fixing each additional peptide bond in the helix is more than compensated by the energy from the new hydrogen bond. In folding the hairpin, since two peptide bonds must be fixed in order to form a single interstrand hydrogen bond, the free energy of hydrogen-bond formation never completely compensates for the loss in conformational entropy. The folded hairpin is therefore unstable unless there are additional favorable side-chain–side-chain interactions. Because more residues must be immobilized before these interactions occur, the barrier for folding the hairpin is significantly larger than for folding the helix, providing an explanation for the slower observed rate. A second important factor in producing faster rates for helices than for hairpins is that a stretch of α-helix may form at any position along the chain with nearly equal probability, whereas a stretch of β-hairpin structure is less probable if it is not centered at the turn. The rate of helix formation is therefore faster, because the effective free-energy barrier is lowered by the larger number of productive ways of forming helical sequences. These two factors provide an explanation for the 20-fold decrease in the rate of folding of the protein G B1 β-hairpin relative to that for the alanine-based α-helix. The negative activation energy for hairpin formation results from the fact that the barrier between the unfolded and folded states is primarily entropic, and the transition state—which is stabilized by the formation of two hydrogen bonds—is lower in enthalpy than the unfolded state.

The folding of a more complex structure—a dimeric coiled-coil formed by a fragment of the GCN4 leucine zipper protein—has also been studied by the Los Alamos group. The native form of the dimer of this 33-residue fragment has the characteristic IR signature of coiled-coils, with a major absorption peak at 1650 cm^{-1}. A relaxation time of approx 100 µs was observed at 333 K ***(27)***. The authors attribute the slow relaxation time to the formation of the coiled-coil tertiary structure. Both the high temperature and the fact that the residues in the coiled-coil sequence have lower helical propensities than alanine combine to make the helical form of the monomeric peptide highly unstable under the conditions of this experiment. Translational entropy also contributes to the barrier for folding, although it is still unclear whether assembly is rate-determining. The low intrinsic stability is compensated by the characteristic network of "zipper" interactions in the folded dimer to produce a stable folded state ***(57)***. The barrier heights are presumably large enough to slow the relaxation rate by a factor of 10^4 relative to that for the formation of alanine helices ***(23,42)***.

4.2. Protein Systems

Nearly all of the laser temperature-jump experiments which have focused on rapid events in the overall folding and unfolding processes have used horse apomyoglobin as a model system. ApoMb exhibits a remarkably complex phase diagram. In addition to the native form (N), which is stable at neutral pH, ApoMb can exist in a number of partially folded states which are thermodynamically accessible at low pH and specific salt conditions. These include an intermediate form (I), which has been shown to be structurally similar to an early folding intermediate, and an extended form (*E*), which is stable at still lower pH. All of these species can be denatured at high temperatures, and the native state can be denatured at low temperatures ***(27,30,58)***. This feature of the phase diagram has been exploited to jump both from folding to unfolding conditions, and from unfolding to folding conditions. All of the folded and partially folded states contain a folded core, consisting of the *A, G,* and *H* helices (the AGH core). These interactions between the C- and N-terminal helices produce species which are all rather compact. The *N, I,* and *E* forms differ in regard to the structure of the remainder of the sequence, which becomes less structured and more extended as the pH is lowered. In the *N* form, the helices are formed and the tertiary structure is similar to that of the holoprotein; in the *I* form the *B, C, D, E,* and *F* helices are formed, but do not exhibit significant tertiary structure; and in the *E* form this portion of the sequence is largely unstructured.

The dynamics of apomyoglobin have been studied using both temperature jumps from the cold-denatured state which induce folding and temperature jumps from the folded states. In the former experiments, folding to *N* final states was monitored at temperatures between 283 K and 293 K by measuring tryptophan fluorescence lifetime data ***(28–30)***. The decay kinetics were found to approach their equilibrium properties in a pair of relaxations—one at 250 ns and one at about 5 μs. The fluorescence lifetime and quantum yield increase in the fast phase and decrease in the slow phase. A spectroscopic investigation of the phase diagram suggested that the initial cold-denatured state in these experiments contains some residual helical structure in the *G* and *H* helices, which is stabilized by helix-helix interactions, while the final *N* state contains the intact *AGH* core ***(58)***. The folding process can therefore be viewed as the formation of a large looped structure, which is stabilized by the specific interactions between the *A* helix and the interacting *GH* helices. The rapid relaxation is interpreted as an increase in the population of unfolded molecules in which the *A* helix is formed, protecting Trp-14, and the slower relaxation as the formation of the *AGH* core in which the fluorescence is quenched by a close interaction with Met-131 in the H-helix. The later assignment is consistent with three additional observations. First, the fluorescence change in the slow phase is abolished by substitution of Met-131 with alanine ***(28)***. Second, the fluorescence lifetime increases in this relaxation if an efficient quencher is added to the solution. Thus, tryptophan becomes buried by other hydrophobic groups and is protected from quenching by the solvent ***(28)***. Third, the lifetime of this relaxation increases in direct proportion to solvent viscosity ***(29)***.

Temperature-jump unfolding studies of the *N* and *E* forms of apomyoglobin in which helix formation is probed by IR spectroscopy have also been carried out at temperatures between 308 K and 320 K ***(25–27)***. Unfolding of the *N* form in D_2O can only be studied at temperatures above 325 K. Temperature jumps on the *N* state produce biphasic kinetics ***(25)***. The fast phase was found to have a relaxation time of about 50 ns at 333 K, while the slow phase had a relaxation time of slightly more than 100 µs. One positive feature of the IR absorption experiments is the ability to distinguish different secondary structural elements by monitoring the kinetics at multiple wavelengths at which the absorption spectra of different structures exhibit peaks. The amplitude of the 50-ns relaxation was greatest at 1632 cm^{-1}, where the A of solvated helices is maximal, while the amplitude of the 120 µs relaxation was largest at 1655 cm^{-1}, where packed helices exhibit maximum A. The equilibrium data suggest that the cooperative melting process being studied is a transition which primarily involves the melting of the folded core residues and all of the change in equilibrium amplitude at 1650 occurs in the 120-µs relaxation. Similar experiments on the *I* form at 303 K showed only a single fast phase with an A change which peaks at 1632 cm^{-1} ***(59)***. This would appear to lead to a straightforward interpretation of the fast relaxation as the rapid equilibration of helices which are not stabilized by any significant tertiary interactions (*C*, *D*, and *F*), and the slow relaxation as arising from melting of the complete folded ensemble, including the core residues. The authors argue, however, that the 120-µs relaxation cannot be assigned to the melting of *AGH* core residues because fluorescence data suggests that the core residues are still folded in the melted form at high temperature ***(60)***. Based on this result, they propose that this relaxation results from the melting of the helices with intermediate stability (*B* and *E*).

A subsequent study of the unfolding of the *E*-form of apomyoglobin, carried out at lower temperatures (308–320 K), showed a similar pair of relaxations ***(26,27)***. A 50-ns process with a spectrum centered at 1630 cm^{-1} and a 60-μs process with a spectrum centered at 1650 cm^{-1} were observed. While the fast process was again attributed to melting of helical residues which are not included in the *AGH* core, the slow process was attributed to the melting of the assembled *AGH* core residues. The weak temperature dependence measured for the slow process—with an apparent activation energy of about 4.5 kcal/mol—comparable to that expected from the temperature-dependence of the solvent viscosity (3.7 kcal/mol), provided additional support for this assignment. The rate of core formation is estimated as about 10^4 s^{-1} using a two-state model.

The relaxation times for core formation measured in the last study (25–100 μs) are remarkably similar to the 40-μs relaxation time observed for formation of the met–65–heme loop after CO photolysis of cytochrome *C* under denaturing conditions ***(16)***, a fact noted by the authors in support of their interpretation. Is it possible that that the folding pathway to both the *N* and *E* forms of apomyoglobin can be viewed as a simple process of diffusion-controlled loop formation? To do so, it would be necessary to rationalize the faster rate for folding from the cold-denatured state as well as the slightly slower rate of folding for the *N* state obtained in the high-temperature melting experiments. Since the *GH* complex is preformed in the low-temperature denatured state, the faster rate may be rationalized by the existence of a larger or more probable target to which the *A*-helix can bind. The much smaller differences between the formation rates for the *N* and *E* forms may arise from any number of factors, including differences in helix propensities and chain stiffness resulting from differences in solution conditions and temperature. Additional studies are clearly required to answer this intriguing question.

5. Conclusion

The laser temperature-jump technique provides a general and powerful method for exploring fast kinetics in folding systems. Although some technical advances in laser hardware can be anticipated—particularly in the development of more robust sources of near-IR light pulses and in pushing the technology to the picosecond time-scale—the fundamental limitations imposed on the production of useful temperature jumps by the physical properties of the solvent may prove to be more restrictive than the currently available laser hardware. The prospects of using other probe techniques—including circular dichroism, UV absorption, Raman spectroscopy, and fourier transform IR spectroscopy ***(61)***—are brighter. As in stopped-flow experiments, the ability to probe the kinetics using a wider variety of spectroscopic signals may significantly improve the ability to assign kinetic processes.

With the number of rapidly-folding small proteins increasing rapidly and the number of designable peptide systems virtually limitless, it is clear that there are wide applications for temperature-jump studies. From this brief discussion, the reader may surmise that it is less difficult to carry out the kinetic experiments than to interpret them unequivocally. Most of the studies summarized here utilize previously characterized systems, or are modest modifications of well-studied systems. Despite this fact, the efforts required to characterize the equilibrium properties of suitable systems for temperature-jump studies and to understand the physics of the spectroscopic probes have frequently outweighed those invested to obtain reliable kinetic data. For novel systems—for which the equilibrium properties have not yet been described—this is an even larger problem. Despite these difficulties, the information gathered from these studies suggests that pushing our knowledge of folding dynamics to rapid process will prove to be enormously useful. The utility stems, in part, from the fact that fast folding processes must necessarily be less complex. This allows them to be modeled using statistical chemical kinetics models, molecular dynamics models, and simplified dynamics models, and these models permit the kinetic details to be more clearly focused than is currently possible with larger and more complex systems.

Acknowledgments

I would like to thank Peggy Thompson, who was responsible for the development of the temperature-jump instrument in our laboratory. Peggy also carried out most of the helix-coil experiments. I would also like to thank my collaborators in applying the temperature-jump technique: Victor Muñoz, who carried out the hairpin experiments; Gouri Jas, who has extended these studies to examine the effects of solvent viscosity; and Steve Hagen, who has used them to look at collapse of cytochrome c. All of the protein folding work in our laboratory at the NIH was carried out in collaboration with William Eaton, who played a major role in kindling my interest in protein folding.

References

1. Bryngelson, J. D. and Wolynes, P. G. (1987) Spin-glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
2. Bryngelson, J. D., Onuchic, J. N., Socci, N.D., and Wolynes, P. G. (1995) Funnels, pathways and the energy landscape of protein folding—a synthesis. *Proteins Struct. Funct. Genet.* **21,** 167–195.
3. Dill, K. A., Bromberg, S., Yue, K., Feibig, K. M., Yee, D. P., Thomas, P. D., and et al. (1995) Principles of protein folding: a perspective from simple exact models. *Protein Sci.* **4,** 561–602.
4. Karplus, M. and Sali, A. (1995) Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **5,** 58–73.
5. Thirumalai, D. and Woodson, S. A. (1996) Kinetics of folding of proteins and RNA. *Accts. Chem. Res.* **29,** 433–439.
6. Shakhnovich, E. I. (1997) Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7,** 29–40.
7. Ptitsin, O. B. (1995) Molten globule and protein folding. *Adv. Prot. Chem.* **47,** 83–229.

8. Sosnick, T. R., Mayne, L. Hiller, R., and Englander, S. W. (1994) The barriers in protein folding. *Nat. Struct. Biol.* **1,** 149–156.
9. Sosnick, T. R., Mayne, L., and Englander, S. W. (1996) Molecular collapse: the rate-limiting step in two-state cytochrome c folding. *Proteins* **24,** 413–426.
10. Colon, W., Elove, G. A., Wakem, L. P., Sherman, F., and Roder, H. (1996) Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry* **35,** 5538–5549.
11. Hosszu, L. L. P., Craven, C. J., Parker, M. J., Lorch, M., Spencer, J., Clark, A. R., et al. (1997) Structure of a kinetic protein folding intermediate by equilibrium amide exchange. *Nat. Struct. Biol.* **4,** 801–804.
12. Jennings, P. A. and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262,** 892–896.
13. Schwarz, G. (1965) On the kinetics of the helix-coil transition of polypeptides in solution. *J. Mol. Biol.* **11,** 64–77.
14. Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G., and Klump, H. (1979) Kinetics of the helix-coil transition of a polypeptide with non-ionic side groups, derived from ultrasonic relaxation measurements. *Biophys. Chem.* **9,** 137–147.
15. Buckler, D. R., Haas, E., and Scheraga, H. A. (1995) Analysis of the structure of ribonuclease A in native and partially denatured states by time-resolved nonradiative dynamic excitation energy transfer between site-specific extrinsic probes. *Biochemistry* **34,** 15,965–15,978.
16. Jones, C. M., Henry, E. R., Hu, Y. Chan, C-K., Luck, S. D., Bhuyan, A., et al. (1993) Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. USA* **90,** 1860–1864.
17. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. USA* **93,** 11,615–11,617.
18. Hagen, S. J., Hofrichter, J., and Eaton, W. A. (1997) Rate of intrachain diffusion of unfolded cytochrome c. *J. Phys. Chem. B* **101,** 2352–2365.
19. Shastry, M. C. R. and Roder, H. (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nat. Struct. Biol.* **5,** 385–392.

20. Shastry, M. C. R., Sauder, J. M., and Roder, H. (1998) Kinetic and structural analysis of submillisecond folding events in cytochrome c. *Acc. Chem. Res.* **31,** 717–725.
21. Thompson, P. A. (1997) Laser temperature jump for the study of early events in protein folding, in *Techniques in Protein Chemistry* (Marshak, D. R., ed.), Academic Press, San Diego, CA, pp. 735–743.
22. Thompson, P. A., Eaton, W. A., and Hofrichter, J. (1997) Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a “kinetic zipper” model. *Biochemistry* **36,** 9200–9210.1.
23. Thompson, P. A., Muñoz, V., Jas, G. S., Henry, E. R., Eaton, W. A., and Hofrichter, J. (2000) The Helix-coil kinetics of an heteropeptide. *J. Phys. Chem. B.* **104,** 378–389.
24. Muñoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of β-hairpin formation. *Nature* **390,** 196–199.
25. Gilmanshin, R., Williams, S. Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1997) Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl. Acad. Sci. USA* **94,** 3709–3713.
26. Gilmanshin, R., Callender, R. H., and Dyer, R. B. (1998) The core of apomyoglobin E-form folds at the diffusion limit. *Nat. Struct. Biol.* **5,** 363–365.
27. Dyer, R. B., Gai, F., and Woodruff, W. H. (1998) Infrared studies of fast events in protein folding. *Acc. Chem. Res.* **31,** 709–716.
28. Ballew, R. M., Sabelko, J., and Gruebele, M. (1996) Observation of distinct nanosecond and microsecond protein folding events. *Nat. Struct. Biol.* **3,** 923–926.
29. Ballew, R. M., Sabelko, J., and Gruebele, M. (1996) Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci. USA* **93,** 5759–5764.
30. Gruebele, M., Sabelko, J., Ballew, R., and Ervin, J. (1998) Laser temperature jump induced protein refolding. *Acc. Chem. Res.* **31,** 699–707.
31. Eigen, M. and de Maeyer, L. (1963) Relaxation Methods in *Techniques in Organic Chemistry, Part II* (Weisberger, A., ed.), Interscience, New York, NY, pp. 895–1054.
32. Flynn, G. W. and Sutin, N. (1974) Kinetic studies of very rapid reactions in solution, in *Chemical and Biochemical Applications of Lasers, C,* (Bradley Moore), ed.), Academic Press, New York, NY, pp. 309–338.

33. Staerk, H. and Czerlinski, G. (1965) Nanosecond heating of aqueous systems by giant laser pulses. *Nature* **205,** 63–64.

34. Beitz, J. V., Flynn, G. W., Turner, D. H., and Sutin, N. (1970) The stimulated Raman effect. A new source of laser temperature-jump heating. *J. Am. Chem. Soc.* **92,** 4130–4132.

35. Turner, D. H., Flynn, G. W., Sutin, N., and Beitz, J. V. (1972) Laser Raman temperature-jump study of the kinetics of the triiodine equilibrium. Relaxation times in the 10^{-8}–10^{-7} second range. *J. Am. Chem. Soc.* **94,** 1554–1559.

36. Ameen, S. (1975) Laser temperature-jump spectrophotometer using stimulated Raman effect in H_2 gas for the study of nanosecond fast chemical relaxation times. *Rev. Sci. Instrum.* **46,** 1209–1215.

37. Turner, D. H. (1986), in *Investigations of Rates and Mechanisms of Reactions* (Bernasconi, C. F., ed.) vol. 6, part 2, Wiley, New York, NY.

38. Williams, A. P., Longfellow, C. E., Freier, S. E., Kierzek, R., and Turner, D. H. (1989) Laser temperature-jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry* **28,** 4283–4291.

39. Holzwarth, J. F., Schmidt, A., Wolff, H., and Volk, R. (1977) Nanosecond temperature-jump technique with an iodine laser. *J. Phys. Chem.* **81,** 2300–2301.

40. Davis, S. J. and Harrington, W. F. (1987) Laser temperature-jump apparatus for the study of force changes in fibers. *Anal. Biochem.* **161,** 543–549.

41. Smith, J. J., McCray, J. A., Hibberd, M. G., and Goldman, Y. E. (1989) Holmium laser temperature-jump apparatus for kinetic studies of muscle contraction. *Rev. Sci. Instrum.* **60,** 231–236.

42. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H., et al. (1996) Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35,** 691–697.

43. Carslaw, H. S. and Jaeger, J. C. (1959) *Conduction of Heat in Solids*, Clarendon Press, Oxford, UK, p. 210.

44. Wray, W. O., Aida, T., Vitols, S. E., and Dyer, R. B. Cavitation and heat transfer effects in laser-induced temperature jump studies of protein folding (unpublished results).

45. Paltauf, G. and Schmidt-Kloiber, H. (1996) Microcavity dynamics during laser-induced spallation of liquids and gels. *Appl. Phys. A.* **62,** 303–311.

46. Frenz, M., Paltauf, G., and Schmidt-Kloiber, H. (1996) Laser-generated cavitation in absorbing liquid induced by acoustic diffraction. *Phys. Rev. Lett.* **76,** 3546–3549.
47. Ballew, R. M., Sabelko, J., Reiner, C., and Gruebele, M. A. (1996) A single-sweep, nanosecond time resolution laser temperature-jump apparatus. *Rev. Sci. Instrum.* **67,** 3694–3699.
48. Blanco, F. J., Jimenez, M. A., Herranz, J., Rico, M., Santoro, J., and Nieto, J. L. (1993) NMR evidence of a short linear peptide that folds into a β-hairpin in aqueous solution. *J. Am. Chem. Soc.* **115,** 5887–5888.
49. Marqusee, S., Robbins, V. H., and Baldwin, R. L. (1989) Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA* **86,** 5286–5290.
50. Scholtz, J. M. and Baldwin, R. L. (1992) The mechanism of alpha helix formation by peptides. *Annu. Rev. Biophys. Biomol. Struct.* **21,** 95–118.
51. Chakrabartty, A. and Baldwin, R. L. (1995) Stability of alpha helices. *Adv. Prot. Chem.* **46,** 141–176.
52. Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G., and Klump, H. (1979) Kinetics of the helix-coil transition of a polypeptide with non-ionic side groups, derived from ultrasonic relaxation measurements. *Biophys. Chem.* **9,** 137–147.
53. Lockhart, D. J. and Kim, P. S. (1993) Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* **260,** 198–202.
54. Lockhart, D. J. and Kim, P. S. (1992) Internal Stark effect measurement of the electric field at the amino terminus of an α helix. *Science* **257,** 947–951.
55. Muñoz, V., Henry, E. R., Hofrichter, J., and Eaton, W. A. (1998) A statistical mechanical model for β-hairpin kinetics. *Proc. Natl. Acad. Sci. USA* **95,** 5872–5879.
56. Jas, G. S., Muñoz, V., Hofrichter, J., and Eaton, W. A. (1999) Effect of viscosity on the kinetics of the helix-coil transition. *Biophys J.* **76,** A175.
57. O’Shea, E. K,, Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. *Science* **254,** 539–544.
58. Sabelko, J., Ervin, J., and Gruebele, M. (1998) Cold-denatured ensemble of apomyoglobin: implications for the early steps of folding. *J. Phys. Chem. B* **102,** 1806–1819.

59. Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H., and Dyer, R. B. (1997) Fast events in protein folding; helix melting and formation in a small peptide. *Biochemistry* **36,** 15,006–15,012.
60. Gilmanshin, R., Dyer, R. B., and Callender, R. H. (1997) Structural heterogeneity of the various forms of apomyoglobin: implications for protein folding. *Protein Sci.* **6,** 2134–2142.
61. Slayton, R. M. and Anfinrud, P. A. (1997) Time-resolved mid-infrared spectroscopy: methods and biological applications. *Curr Opin. Struct. Biol.* **7,** 717–721.

9

Kinetics of Conformational Fluctuations by EX1 Hydrogen Exchange in Native Proteins

T. Sivaraman and Andrew D. Robertson

1. Introduction

Nearly half a century ago, Hvidt and Linderstrøm-Lang established amide hydrogen (NH) isotope exchange as a fruitful probe of structure and dynamics in proteins ***(1–4)***. Rapid growth in the use of NH exchange over the last 15 yr has followed from advances in both technology (e.g., multidimensional nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry (MS) and rapid-mixing methodologies) and in our understanding of how NH exchange in proteins is slowed relative to exchange in small peptides ***(5–11)***. Most applications of NH exchange involve one of the two major types of exchange experiments. These are distinguished by the population of native protein: the first and most popular type of experiment involves native protein alone, and the second type is used to follow exchange during the process of folding or unfolding. For practical reasons, almost all exchange experiments in proteins focus on the NHs of the peptide backbone. This chapter focuses on the exchange of these NHs from the native protein.

The appeal of hydrogen exchange is fourfold. First, the experiments are relatively easy to execute. In a typical hydrogen/deuterium

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

(H/D) exchange experiment, protein in H_2O is dissolved or diluted into D_2O and the replacement of protein hydrogens by solvent deuterons is monitored by NMR or MS. Second, NMR and MS can be used to monitor exchange simultaneously at many individual residues in proteins. Third, the replacement of hydrogen by deuterium has little or no effect on protein conformation and very modest effects on protein stability. Fourth, a simple kinetic model for slow exchange in native proteins has proven to be accurate, and thus provides a robust framework for interpreting exchange data ***(4)***. This model is described in more detail in **Subheading 2.**, but the principles of interpretation are introduced here to illustrate the application of exchange to studies of protein stability and protein folding.

Investigators typically execute exchange experiments under solution conditions favoring one of two kinetic limits to the model. In the EX1 limit, the rate of H/D exchange directly reflects the rate at which individual NHs are exposed to solvent, whereas the equilibrium constant of the opening reaction dictates the rate of exchange in the EX2 limit. In general, EX2-type exchange is a powerful method to estimate the relative stability of various regions of protein molecules. This in turn sheds light on the structural basis for the thermodynamic stability of proteins at residue-level resolution (*see* **refs. *12–21*** and Chapter 4). EX1-type exchange has received less attention over the years, but recent developments suggest that EX1-type exchange may provide useful information regarding the kinetics of conformational interconversions. These may occur within the native state and between the native and denatured states, even under solution conditions favoring the native state.

Some investigators have recently proposed that rates of H/D exchange in native proteins may be used to predict the overall chronology of events that occur during protein folding (***18***,***22***, and ***23***). However, this is a subject of heated debate ***(24–27)***. In this regard, experimental investigation of both EX2 and EX1 exchange kinetics in native proteins may provide useful insight regarding this controversy ***(28,29)***. For example, EX1-type exchange has recently been used to investigate the kinetics of solvent exposure for nearly every hydrogen-bonded NH of turkey ovomucoid third domain (OMTKY3) under native conditions ***(30–33)***. In principle, these types of studies could be combined with studies of the kinetics of folding and unfolding to determine if the order of events detected in native protein is the same. This is just one example of the potential application of EX1-type exchange. Overall, EX1 exchange probably provides more direct experimental access to the dynamics of hydrogen-bond interactions in native proteins than EX2-type exchange. This follows from the fact that rate constants for EX1-type exchange are direct measures of conformational kinetics, while EX2-type exchange relies entirely on *a priori* knowledge of the chemistry of exchange under the experimental solution conditions. Generally, the most successful approach typically involves a combined analysis of EX2- and EX1-type exchange.

The major goal of this chapter is to outline the identification and applicability of EX1-type exchange in native proteins. This chapter begins with a brief discussion of the basic principles of the H/D exchange reaction in proteins. **Subheading 3.** includes two types of experiments used by our laboratory to examine EX1-type exchange in native proteins.

2. Mechanism of H/D Exchange in Proteins

2.1. Kinetic Mechanism for Exchange

Substitution of solvent deuterons for protein-bound protons or vice versa is commonly denoted as H/D exchange. Protons attached to nitrogen, oxygen, and sulfur atoms are labile to exchange with solvent hydrogens. In studying the conformational motility of proteins, peptide backbone NHs are generally the most useful probes in H/D exchange. This follows from the fact that most backbone NHs are involved in hydrogen-bonded structure, which appears to be the major determinant of slow NH exchange in proteins. Slow exchange of hydrogen-bonded and buried NHs is generally interpreted with the two-state model proposed by Hvidt ***(4)***.

$$NH_{(C)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} NH_{(O)} \xrightarrow{k_{rc}} ND \qquad (1)$$

According to this model, slowly-exchanging NHs can occupy only two possible conformations: a closed state, $NH_{(C)}$, and an open state, $NH_{(O)}$. The rate constants k_{op} and k_{cl} are for the opening and closing reactions, respectively. Exchange takes place only from the open state, with the rate constant denoted k_{rc}. This is the rate constant for exchange from a random coil conformation, and is governed by many factors such as protein sequence, pH, temperature, and denaturant concentration. The value of k_{rc} can be predicted using the method of Bai et al. ***(34)***.

According to the model of scheme 1, the observed exchange-rate constant for slow NH exchange, k_{ex}, can be defined as:

$$k_{ex} = (k_{op} \cdot k_{rc})/(k_{op} + k_{cl} + k_{rc}) \qquad (2)$$

For native proteins, one can usually assume that $k_{op} << k_{cl}$, in which case **Eq. 2** simplifies as follows:

$$k_{ex} = (k_{op} \cdot k_{rc})/(k_{cl} + k_{rc}) \qquad (3)$$

The two kinetic limits discussed in **Subheading 1.** can be derived from **Eq. 3**. If $k_{cl} >> k_{rc}$, then k_{ex} can be expressed as:

$$k_{ex} = (k_{op}/k_{cl}) \cdot k_{rc} = K_{eq} \cdot k_{rc} \qquad (4)$$

where K_{eq} is the equilibrium constant for the opening reaction. This kinetic limit is known as EX2-type exchange (pseudo-bimolecular reaction). In the case where $k_{rc} >> k_{cl}$, then k_{ex} will equal k_{op}.

$$k_{ex} = k_{op} \qquad (5)$$

This kinetic limit is known as EX1-type exchange (pseudo-unimolecular reaction).

For both EX2- and EX1-type exchange, apparent first-order kinetics are expected. The value of k_{ex} can thus be derived by fitting the experimental proton occupancy as a function of time to a simple exponential function.

Exchange data for a given NH are often reported as a protection factor, P:

$$P = k_{rc}/k_{ex} = k_{cl}/k_{op} = 1/K_{eq} \tag{6}$$

Under EX2 conditions, the protection factor is related to the apparent free energy of exchange, ΔG_{ex}, as shown in **Eq. 7**:

$$\Delta G_{ex} = -RT \ln (1/P) \tag{7}$$

where R is the gas constant and T is the absolute temperature.

2.2. Molecular Mechanisms for Exchange

The accuracy of the simple two-state model for slow NH exchange in native proteins has been tested and verified. These tests have taken advantage of the fact that the most slowly-exchanging NHs in native proteins appear to exchange only upon complete unfolding (*see* **ref. *35*** and Chapter 4). For these NHs, ΔG_{HX} is usually equal to the free energy of unfolding, $\Delta G_u°$, over a wide range of conditions. These NHs tend to be involved in secondary structure and buried in the protein interior.

The amide protons that show more rapid-exchange kinetics are typically involved in hydrogen-bond interactions at the periphery of secondary structure and in loops. For these NHs, ΔG_{HX} is less than $\Delta G_u°$. The molecular events that are responsible for exchange at these NHs are unknown, but they are usually referred to as “local” fluctuations to distinguish them from the global unfolding reaction that is responsible for the most slowly-exchanging NHs.

The contribution of global unfolding and local fluctuations to slow NH exchange can be detected by examining the exchange reaction as a function of denaturant under EX2 conditions (*see* **refs. *36*,*37*** and Chapter 4). Global unfolding is generally much more sensitive than local fluctuations to increasing denaturant concentrations. Destabilization by chemical denaturant—or by any other perturbation such as temperature or pH—leads to an decrease in $\Delta G_u°$. For NHs that exchange by global unfolding, this decrease is mirrored precisely in ΔG_{HX} *(18)*. In contrast, ΔG_{HX} values for NHs that exchange by local fluctuations are largely indifferent to decreases in $\Delta G_u°$—at least until it overtakes ΔG_{HX}. At this point, global unfolding also becomes the dominant contributor to slow exchange at these sites, and consequently, $\Delta G_{HX} = \Delta G_u°$ *(16)*.

The appeal of EX1-type exchange relative to EX2-type exchange is that it provides relatively direct information regarding the kinetics of conformational motion at individual residues in native proteins (**Eq. 5**). Systematic investigation of EX1-type exchange is only now being conducted, so the information gathered from these studies is likely to increase steadily over the next few years as we gain more insight into the molecular events that are responsible for slow NH exchange.

3. Methods

3.1. Strategies for Promoting EX1-Type Exchange

As discussed in **Subheading 2.**, observation of EX1 vs EX2 exchange kinetics depends solely on the rate constants for the closing, k_{cl}, and intrinsic exchange reactions, k_{rc} (**Eqs. 4,5**). EX1 exchange occurs when $k_{rc} >> k_{cl}$, and EX2 is observed when $k_{cl} >> k_{rc}$. For example, k_{cl} for the most slowly-exchanging NHs—those that exchange upon complete unfolding—probably corresponds to the rate constant of folding. For small globular proteins, this may be on the order of 1000 s^{-1} *(38)*. Values of k_{rc} only exceed 1000 s^{-1} at pH >9, so alkaline pH will be needed for the observation of EX1-type exchange in such cases. For the more rapidly-exchanging NHs, k_{cl} values are likely to be greater, so even more alkaline pH conditions will be necessary for observation of EX1-type exchange at these sites.

In general, the two basic approaches to observing EX-1-type exchange are to either accelerate k_{rc} or decrease k_{cl}. pH and temperature can be used to accelerate k_{rc}. At pH >3, k_{rc} is dominated by specific base catalysis, so k_{rc} increases by a factor of 10 with every unit increase in pH. Increasing temperature by 10˚C leads to an approximately threefold increase in k_{rc} ***(34)***. Temperature, pH, and chemical denaturants may be used to reduce k_{cl}. For the most slowly-exchanging NHs, perturbations that reduce protein stability are likely to do so by decreasing the rate constant for folding, which is k_{cl} for these NHs. On the other hand, one should keep in mind that these perturbations will also probably increase k_{op} for these NHs.

This chapter has focused explicitly or implicitly on the role of protein stability in optimizing conditions for the observation of EX1-type exchange in native proteins. Any attempt to detect EX1-type exchange is thus best preceded by an investigation of protein stability under the solution conditions in which exchange is to be investigated. In the following sections, varying alkaline pH is the principal means for detecting EX1-type exchange, so protein stability should be determined over this range of pH using thermal or chemical denaturation (*see* Chapters 2 and 3). D_2O typically has small but significant effects on protein stability ***(39)***, so stability studies should be carried out in this solvent.

3.2. Identifying EX1-Type Exchange

How does one distinguish between EX1- and EX2-type exchange experimentally? In favorable cases, the simplest diagnostic is the pH dependence of exchange. The observed rate constant—k_{ex}—for EX2-type exchange—includes k_{rc}, so the pH dependence of k_{rc} will also be manifested in the pH dependence of EX2-type exchange (**Eq. 4**). In contrast, k_{ex} for EX1-type exchange has no explicit pH dependence (**Eq. 5**). One thus detects a switch from EX2- to EX1-type exchange with increasing pH by the disappearance of a pH dependence for NH exchange (*see* **Fig. 1**). In reality, one must be aware of the possibility that k_{op} and k_{cl} may also be dependent on pH, and the stability studies discussed previously may provide useful information in this regard.

Another possible experimental indication of a switch from EX2- to EX1-type exchange is anomalously large values of ΔG_{HX} ***(40)***. This approach combines data obtained under EX2 conditions, knowledge of stability under the solution conditions for H/D

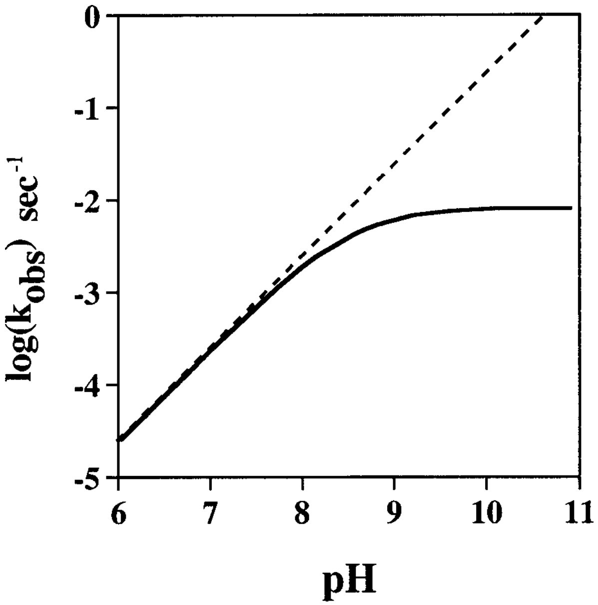


Fig. 1. Hypothetical pH dependence of exchange depicting a switch from EX2 to EX1 exchange (*solid line*) and pure EX2 exchange (*dashed line*) for slowly exchanging amide protons. Values of k_{ex} were simulated using **Eq. 3** and 0.008 s^{-1} and 2000 s^{-1} for k_{op} and k_{cl}, respectively. Values of k_{rc} were calculated using a second-order rate constant for base-catalyzed exchange of $2.6 \times 10^9 M^{-1}s^{-1}$ ***(34)***.

exchange, and investigation of exchange as a function of more alkaline pH. Data for stability and exchange under EX2 conditions—such as at pH <7—provides information on the relationship between values of ΔG_{HX} and $\Delta G_u°$. Based on these data, knowledge of $\Delta G_u°$ at more alkaline pH is used to predict ΔG_{HX}—and consequently, k_{ex} under these conditions—assuming that exchange is still in the EX2 limit. We call this the "EX2 hypothesis" ***(32)***. If the observed k_{ex} at alkaline pH is less than the predicted k_{ex}, i.e., the observed ΔG_{HX} exceeds the predicted value—then the simplest explanation is that the EX2 hypothesis is incorrect and one is detecting EX1-type exchange.

Two other methods to detect EX1-type exchange rely on correlated H/D exchange at multiple NHs ***(7,28)***. Under EX1 conditions, the rate-limiting step in H/D exchange is an opening reaction. If multiple slowly-exchanging NHs are exposed to solvent in the same opening reaction, then these NHs will undergo correlated exchange—they will all exchange at the same time. For two NHs that are within 5 Å of one another, one may be able to use NMR and the nuclear Overhauser effect (NOE) to determine if H/D exchange at these NHs is correlated ***(28)***. For example, if H/D exchange is correlated, then the magnitude of the NOE will remain constant relative to the NH peak intensities. If, however, exchange is uncorrelated, then the NOE intensity will decay more rapidly than the NH peak intensities. This method may be technically challenging because one is trying to detect relatively small differences in the intensities of small NMR peaks.

Electrospray-ionization MS (ESI-MS) can also be used to detect correlated H/D exchange ***(7,31)***. In this case, correlated exchange results in broadening or splitting of MS peaks. The number of mass units represented in the broadening or splitting provides an estimate for the number of NHs involved in the correlated exchange event. One should recall that both the MS and NOE methods for detecting EX1-type exchange rely on the assumption that EX1-type exchange necessarily leads to correlated exchange. However, uncorrelated or partially correlated EX1-type exchange has been observed ***(31)***. Thus, positive results for the MS and NOE experiments are indicative of EX1-type exchange, but negative results are not proof of EX2-type exchange.

Two different methods for using NMR to measure H/D exchange at alkaline pH are presented in **Subheadings 3.3.** and **3.4.** Alkaline pH provides the most straightforward condition for observation of EX1-type exchange, but the approach is predicated on reasonable protein stability at alkaline pH. Ample precedent for alkaline pH stability is found in enzymology, where alkaline pH has been used for many years in probing acid/base catalysis ***(41)***.

The two methods for monitoring H/D exchange at alkaline pH are distinguished by the kinetics of exchange. In cases where the k_{ex} is less than about 0.01 s^{-1}, one can measure H/D exchange in real-time (*see* **Subheading 3.2.**). Residue-specific values for k_{op} and k_{cl} constants are estimated by examining the pH dependence of k_{ex} ***(30,31)***. In our laboratory, these conditions have been met by the most slowly exchanging NHs in OMTKY3 and ubiquitin. In contrast, the more rapidly-exchanging NHs exchange completely in the dead time of such experiments. Consequently, the quenching strategy described in **Subheading 3.4.** is used for estimating residue-specific values of k_{op} and k_{cl} for these NHs, where local structural fluctuations rather than global unfolding appear to dominate exchange.

3.3. EX1-Type Exchange for the Most Slowly-Exchanging NHs

3.3.1. Experimental Procedure

In some cases, EX1-type exchange may be detected by monitoring the pH dependence of H/D exchange in real-time using ^{1}H NMR spectroscopy ***(30)***. For the following protocol, H/D exchange is initiated by dissolving lyophilized protein in D_2O. For proteins that do not lyophilize well, exchange may be initiated by diluting protein in H_2O into a relatively large vol of D_2O. A typical experimental procedure follows:

1. Protein concentration should be optimized, taking into account solubility and the sensitivity needs of the experiment. H/D exchange may be followed by NMR in a number of different ways, including one-dimensional (1D) ^{1}H NMR, two-dimensional (2D) ^{1}H-^{1}H experiments such as correlation spectroscopy (COSY) or total correlation spectroscopy (TOCSY), and 2D ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) with ^{15}N-enriched proteins. Typically, protein concentrations for 1D and 2D ^{1}H NMR experiments are $\geq$ 0.5 m*M* and 2 m*M*, respectively, for data acquisition on a 500-MHz NMR spectrometer. Measurement of very rapid exchange may require higher protein concentrations in order to acquire NMR data in a shorter time. With ^{15}N-labeled protein, the concentration required for an HSQC experiment can be as low as 0.1 m*M*. 1D data may be acquired in as little as a few seconds if protein concentration is high and little or no signal averaging is needed, but a more typical

acquisition time may be 30–60 s. Relative to 1D NMR data, 2D ^{1}H spectra typically take at least 25 times as long to acquire at the same protein concentration. In general, the best approach is to use ^{1}H-^{15}N HSQC with pulsed-field gradients, if possible ***(42,43)***. Under these circumstances, good 2D spectra can be acquired in as little as 1–2 min.

2. Appropriate buffer solutions should be carefully chosen to follow the exchange reaction at the desired pH. Two major factors should be considered: buffer strength and spectral interference. The pK_a of the buffer should be as close to the desired pH as possible. With regard to spectral interference, ^{1}H resonances from the buffer may obscure protein NH resonances. Moreover, even in cases where direct overlap is not a problem, large buffer signals are likely to degrade the spectral quality. Hence, deuterated or nonprotonated buffers, such as phosphate or borate, should be used when possible. The influence of buffer on protein stability should be assessed.
3. In initiating the H/D exchange experiment, the protein sample should be preadjusted to the desired pH prior to lyophilization. The sample should be lyophilized to constant weight to minimize residual H_2O.
4. All NMR data-acquisition parameters should be preset using a mock sample to minimize the dead-time, which is the time between initiation of exchange and the start of data acquisition. This sample can also be used to optimize the data-acquisition parameters to achieve the best compromise between rapid data accumulation and sensitivity. The experimental temperature can be calibrated using methanol (180–300 K) or ethylene glycol (300–420 K) standards ***(44)***.
5. To minimize the time required for the sample to reach the reaction temperature, the NMR tube, protein sample, and buffer solution are equilibrated at the experimental temperature prior to initiating H/D exchange.
6. H/D exchange is initiated by dissolving dry protein in buffered D_2O. The protein sample is rapidly transferred to the NMR tube, and data acquisition is started in as little as 1–2 min. To minimize the dead-time, sample pH is measured after the experiment is complete. This measurement is essential: preadjustment of buffer and protein pH assures that the experimental pH will be close to the desired pH, but precise knowledge of the experimental pH only comes from direct measurement.
7. H/D exchange kinetics are examined as a function of time at various pH conditions. For data acquisition at a given pH, the time interval

between NMR spectra should be sufficient to sample the decay of ^1H intensity at a minimum of five time points. Sampling at more time points will yield more precise and accurate rate constants.

3.3.2. Data Analysis

Identical processing parameters should be used for all NMR spectra in a set. Careful attention should be given to proper phasing and baseline adjustments in the transformed spectra. NH resonance intensities can be estimated by measuring peak height in one-dimensional ^1H NMR and peak height or peak vol in two-dimensional NMR spectra. Ideally, NH intensities should be internally referenced to nonexchangeable resonances, such as aromatic protons or aliphatic protons in ^1H spectra. To determine rate constants for exchange, NH resonance intensity (*I*) as a function of time (*t*) is fit to a single exponential function:

$$I = I_o \exp(-k_{ex}t) + C, \tag{8}$$

where I_o is initial intensity, k_{ex} is the rate constant of exchange, and *C* is final amplitude. This fitting can be performed using commercially available programs that are equiped for nonlinear least squares analysis.

The residue-specific k_{op} and k_{cl} can be estimated by fitting the pH dependence of k_{ex} to the following modified version of **Eq. 3**:

$$k_{ex} = (k_{op}*k_b*10^{-(pK_D - pD)}) / (k_{cl}+k_b*10^{-(pK_D - pD)}) \tag{9}$$

in which pK_D is the dissociation constant of D_2O and pD is pH + 0.4, where pH is the reading on a glass electrode. k_b is the predicted second-order rate constant for base-catalyzed exchange ***(34)***.

The fits are likely to converge successfully if k_{ex} values are $\geq 0.5 \cdot k_{op}$ at the highest pH. On a plot of log k_{ex} vs pH, such as that shown in **Fig. 1**, this is the point at which the slope is 0.5. Qualitatively, the data on such plots should show a systematic decrease in k_{ex} relative to values expected for EX2-type exchange, as indicated by the straight line in **Fig. 1**.

3.4. EX1-Type Exchange for More Rapidly-Exchanging NHs

The majority of slowly-exchanging NHs in a native protein probably exchange too rapidly to detect by real-time measurements of H/D exchange at alkaline pH. For these NHs, some type of manual rapid-mixing or quench-flow type of experiment will be needed. Manual mixing and quench-flow experiments can be used to determine k_{ex} values as large as approx 0.5 and 500 s^{-1}, respectively. The quench-flow experiments are performed on specialized instrumentation which is available from a number of commercial sources. The general principles of quench-flow experiments are described by Fersht ***(41)***.

Alternatively, saturation- or magnetization-transfer experiments are feasible in cases where k_{ex} values fall between approx 0.1 and 10 s^{-1} ***(43,45)***. In these experiments, protein is dissolved in 90% H_2O/10% D_2O (v/v), and NH exchange with solvent is monitored by perturbations of NH peak intensities when the solvent resonance is saturated with a radio frequency pulse. In contrast to the quench-flow experiments, saturation-transfer requires only a single protein sample for determining k_{ex} at a given solution condition. Thus, saturation-transfer experiments should be pursued if possible. In fact, these are useful preliminary experiments, as described here, even when quench-flow experiments are ultimately needed for accurate measurements of k_{ex} values.

The rapid-mixing or quench-flow approach takes advantage of the sensitivity of H/D exchange to pH—values of k_{rc} at pH 3.0 are decreased by six orders of magnitude relative those at pH 9.0. The general protocol thus involves initiating exchange through rapid addition of alkaline-buffered D_2O to a protein solution, then exchange for a known time at alkaline pH, and finally quenching of further exchange by rapid acidification. The simplest quench-flow experiment uses three syringes and two rapid mixers ***(41)***. Syringes 1 and 2 are connected by mixer 1, which is followed by a delay line whose known vol is under the control of the experimentalist. This delay line terminates at mixer 2, which is also connected to syringe 3. For an H/D exchange experiment, syringe 1 contains protein solution, syringe 2 contains alkaline-buffered D_2O, and syringe 3 contains acid-buffered D_2O. During the experiment, all three syringes are pushed simultaneously, and the total time of H/D exchange at alkaline pH, t_{ex}, is determined by the speed of the push, v_{push} in $ml \cdot s^{-1}$, and the vol of the delay line, vol in mL: i.e., $t_{ex} = (vol/v_{push})$ s.

In principle, the approach outlined in **Subheading 3.3.** can be modified to include this quenching protocol. However, a separate protein sample will now be required for each time point. For exchange at a given pH, one is thus likely to consume at least ten times as much protein as in a real-time experiment. This is the best procedure, if possible, but the constant-time approach described here provides a reasonable alternative that consumes much less protein ***(32,33)***.

A key feature of the constant-time approach is the use of a very short and constant time period for exchange at various pH values. One thus monitors NH peak amplitude as a function of pH. In the event of EX1-type exchange, NH peak intensities show decreased pH dependences relative to those expected for EX2-type exchange. **Eq. 10** describes NH peak intensity, *I*, as a function of pH and with a constant t_{ex}:

$$I = I_0 \exp\left(\frac{-t_{ex} k_{op} k_b 10^{-(pK_D - pH)}}{k_{cl} + k_b 10^{-(pK_D - pH)}}\right) \qquad (10)$$

where all of the parameters are as described for **Eq. 9**.

The basic logic of monitoring the amplitude at a single time-point can perhaps be best understood by inspection of **Fig. 2**. The simulations in **Fig. 2A** and **2B** have been generated using k_{op} and k_{cl} values of 140 s^{-1} and 2×10^4 s^{-1}, respectively, and a fixed exchange time of 12 ms. **Fig. 2A** depicts the time course of H/D exchange at various pHs; the vertical line corresponds to 12 ms of exchange. At lower pH values, where EX2-type exchange dominates, the rate of exchange is accelerated as a function of base concentration. As the EX1 limit is approached at alkaline pH, the rate acceleration by pH

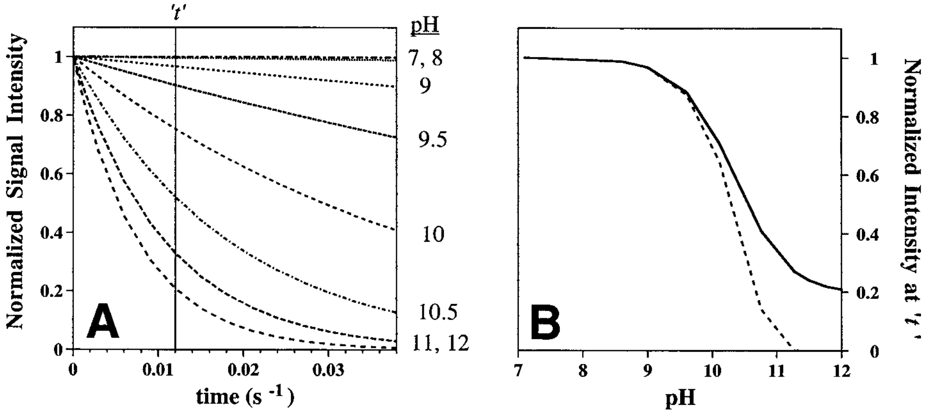


Fig. 2. Kinetic decay curves simulated at various pHs. **(A)** Simulations were performed using **Eq. 10** and 140 s^{-1} and 20,000 s^{-1} and $1.9 \cdot 10^8\ M^{-1}s^{-1}$ for k_{op}, k_{cl}, and k_b respectively. A vertical line is drawn at t_{ex} = 12 ms. **(B)** Simulated curves of the pH dependence of NH peak intensity at t_{ex} = 12 ms. The solid curve depicts a switch from EX2 to EX1 exchange at a very alkaline pH, while the dashed curve indicates the expected amplitudes for pure EX2 exchange.

tapers off. The solid curve of **Fig. 2B** shows the peak intensity at 12 ms vs pH, and the dashed curve shows the expected pH dependence for EX2-type exchange. To detect an approach to EX1-type exchange, one thus looks for NH peak intensities that are systematically greater than those expected for EX2-type exchange.

3.4.1. Experimental Procedure

1. Identification of an appropriate fixed time and temperature for exchange are key steps. To this end, we use fairly crude saturation-transfer experiments in which 1D ^1H NMR data are collected on protein in H_2O and at various temperatures and alkaline pH values. For these experiments, it is useful to have estimates for the spin-lattice relaxation time constants—T_1 values—for the NH resonances. Rough estimates are sufficient for initial studies, and these can be readily obtained from inversion- or saturation-recovery experiments using 2D ^1H NMR ***(46)***. Typical T_1 values for NH reso-

nances of OMTKY3 and ubiquitin are about 0.5 s, and these values are not very temperature-sensitive. For estimation of k_{ex} using the saturation-transfer experiment, the H_2O resonance is suppressed by irradiation with a low-power pulse of relatively long duration, ≥2 s. Intensities for cross-peaks involving NH resonances are monitored as a function of pH. The relationship between peak intensity, I, and k_{ex} is:

$$I/I_0 = 1/(1 + k_{ex}T_1) \tag{11}$$

where I_0 is NH peak intensity in the absence of exchange; this can be measured at more acidic pH values. For OMTKY3 at 30°C and pH >10, the signals for the most slowly exchanging NHs are visible in the 2D NMR spectrum, but those for the more rapidly exchanging NHs are not. If one assumes a value of 0.1 for the minimum detectable peak intensity, then these results suggest that k_{ex} for these NHs is $>20\ s^{-1}$. Quench-flow experiments are needed to determine k_{ex} for these NHs at pH >10.

2. For quench-flow studies of OMTKY3, the quench-flow device is equilibrated at the desired temperature with a circulating water bath prior to loading the syringes.
3. Syringe 1 is filled with a freshly prepared 5-m*M* protein soln in D_2O containing 0.1 *M* potassium phosphate, 0.1 *M* perdeuterated glycine (Cambridge Isotope Laboratories), and 25 m*M* sodium tetraborate adjusted to pH 3.0 with concentrated DCl. The sample is prepared and stored on ice until loaded into syringe 1. The rationale for dissolving protein in D_2O is that this reduces the overall vol of D_2O needed for the experiment and the extent of dilution needed to initiate H/D exchange at alkaline pH. This helps to minimize the concentration of protein in syringe 1. At pH 3.0 and on ice, the H/D exchange reaction is very slow, even for solvent-exposed NHs, so little or no exchange occurs prior to the experiment.
4. Syringe 2 is filled with D_2O containing 0.77 *M* NaOD, and syringe 3 is filled with D_2O containing 0.27 *M* DCl.
5. The desired pH values for alkaline H/D exchange are determined prior to the experiment by titration of the syringe 1 solution with the syringe 2 solution. The volume of syringe 3 solution needed for the acid quench is likewise determined by titration of the mixture of syringe 1 and 2 solutions with syringe 3 solution. Typical volumes

for studies of OMTKY3 are 200 µL of syringe 1 soln, 50–160 µL of syringe 2 soln to cover pH ranging from 7–12, and 250 to 350 µL of syringe 3 soln.

6. One possible concern is accurate determination of very alkaline pH values, especially in the presence of sodium ion ***(47)***. The pH electrode should be calibrated at two pH values that bracket the desired pH values. Commercial standards are available up to pH 11.0. A degassed and saturated solution of calcium hydroxide provides a relatively convenient standard at about pH 12.2; precise pH values as a function of temperature are reported in the text by Bates ***(48)***. The magnitude of the alkaline sodium effect depends on the type of electrode, and is often reported by the manufacturer. For good-quality glass electrodes, this effect typically leads to underestimation of the true pH by no more than approx 0.1 pH units at pH 12.0.
7. After the syringes have been filled, some of the syringe solutions are passed through the mixers and delay line in “mock” experiments in order to establish the proper initial conditions for the experiment. This necessarily entails the sacrifice of protein solution. In general, the investigator should be prepared to lose about half of the protein solution to these washing steps.
8. The experiments are then executed by pushing known volumes from syringes 1 and 2 through the mixer delay line at a known flow rate, v_{push} in $ml \cdot s^{-1}$. For studies of OMTKY3, v_{push} is 2 $mL \cdot s^{-1}$. This flow rate should be rapid enough to ensure complete mixing. The instrument manufacturer should be consulted regarding the relationship between flow rate and the efficiency of mixing. For OMTKY3, the vol of the delay line is 24 µL, yielding a t_{ex} of 12 ms.
9. The pH of alkaline H/D exchange is systematically varied by adjusting the volume of syringe 2 solution, as determined by titration in **step 5**. The mixers and delay line usually need to be flushed between pH points, as described in **step 7**. To optimize subsequent least-squares fitting of the data, the pH range should be sufficient to describe both a region of full NH peak intensity at lower pH and significant changes in peak intensity at more alkaline pH (*see* **Fig. 2B**). For OMTKY3, H/D exchange was measured at about 10.0 values ranging from pH 7.0–pH 12.2.
10. The resulting samples contain about 1.5 m*M* OMTKY3 in 0.5 mL. These samples should be stored frozen or on ice. In subsequent steps, all samples should be handled identically as much as possible.

11. Samples are transferred to NMR tubes for acquisition of 2D NMR data. For OMTKY3, TOCSY experiments are used. Each two-dimensional data set is acquired in about 2 h, and consists of 256 blocks of t_1 data. Each block is the sum of eight transients digitized in 2048 data points. The residual solvent peak is suppressed by irradiation of the water peak.
12. A control sample representing complete H/D exchange should be prepared by manual mixing of the appropriate volumes of syringe 1 and syringe 2 solutions. This is followed by addition of the appropriate vol of syringe 3 solution. This control sample is essential for proper quantitative interpretation of NH peak intensities vs pH; some residual NH peak intensity is to be expected after complete H/D exchange because of residual H_2O present during the alkaline pH exchange.

3.4.2. Data Analysis

Data analysis largely parallels that described in **Subheading 3.3.2.** The constant-time data are fit to **Eq. 10**. For the fits to converge, the data must describe a low-pH plateau of full NH peak intensity, a region of declining intensity at intermediate pH values, and a plateau or approach to a plateau of nonzero intensity at high pH (*see* **Fig. 2B**). In this regard, simulations of the EX2 hypothesis are a useful guide for determining whether or not the data indicate EX1-type exchange.

Acknowledgment

We thank Dr. Cammon B. Arrington for helpful discussions on the manuscript. This research was supported by a grant from the NIH (GM46869).

References

1. Hvidt, A. and Linderstrøm-Lang, K. (1955) The pH-dependence of the deuterium exchange of insulin. *Biochem. Biophys. Acta* **18,** 308.
2. Hvidt, A. (1955) Deuterium exchange between ribonuclease and water. *Biochim. Biophys. Acta* **18,** 307–308.

3. Hvidt, A. and Nielsen, S. O. (1966) Hydrogen exchange in proteins. *Adv. Protein Chem.* **21,** 287–386.
4. Hvidt, A. (1964) A discussion of the pH dependence of the hydrogen-deuterium exchange in proteins. *CR Trav. Lab. Carlsberg* **34,** 299–317.
5. Udgaonkar, J. B. and Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335,** 694–699.
6. Roder, H., Elove, G. A., and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and proton NMR. *Nature* **335,** 700–704.
7. Miranker, A., Robinson, C. L., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993) Detection of transient protein folding populations by mass spectrometry. *Science* **262,** 896–900.
8. Kiefhaber, T. and Baldwin, R. L. (1995) Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proc. Natl. Acad. Sci. USA* **92,** 2657–2661.
9. Gladwin, S. T. and Evans, P. A. (1996) Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labeling. *Folding and Design* **1,** 407–417.
10. Laurents, D. V., Bruix, M., Jamin, M., and Baldwin, R. L. (1998) A pulse-chase-competition experiment to determine if a folding intermediate is on or off-pathway: application to ribonuclease A. *J. Mol. Biol.* **283,** 669–678.
11. Robertson, A. D. and Baldwin, R. L. (1991) Hydrogen exchange in thermally denatured ribonuclease A. *Biochemistry* **30,** 9907–9914.
12. Woodward, C. K. and Hilton, B. D. (1979) Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu. Rev. Biophys. Bioeng.* **8,** 99–127.
13. Roder, H. (1986) Structural characterization of protein folding intermediates by protein magnetic resonance and hydrogen exchange. *Meth. Enzymol.* **176,** 446–473.
14. Mayo, S. L. and Baldwin, R. L. (1993) Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* **262,** 873–876.
15. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* **269,** 192–197.

16. Swint-Kruse, L. and Robertson, A. D. (1996) Temperature and pH dependences of hydrogen exchange and global stability for ovomucoid third domain. *Biochemistry* **35,** 171–180.
17. Hosszu, L. L., Craven, C. J., Lorch, M., Spencer, J., Clarke, A. R., and Waltho, J. P. (1997) Structure of a kinetic protein folding intermediate by equilibrium amide exchange. *Nat. Struct. Biol.* **4,** 801–804.
18. Xu, Y., Mayne, L., and Englander, S. W. (1998) Evidence for an unfolding and refolding pathway in cytochrome c. *Nat. Struct. Biol.* **5,** 774–778.
19. Sivaraman, T., Kumar, T. K. S., and Yu, C. (1999) Investigation of the structural stability of cardiotoxin analogue III from the Taiwan cobra by hydrogen-deuterium exchange kinetics. *Biochemistry* **38,** 9899–9905.
20. Chakshusmathi, G., Ratnaparkhi, G. S., Madhu, P. K., and Varadarajan, R. (1999) Native-state hydrogen-exchange studies of a fragment complex can provide structural information about the isolated fragments. *Proc. Natl. Acad. Sci. USA* **96,** 7899–7904.
21. Neira, J. L., Selvilla, P., Menendez, M., Bruix, M., and Rico, M. (1999) Hydrogen exchange in ribonuclease A and ribonuclease S: evidence for residual structure in the unfolded state under native conditions. *J. Mol. Biol.* **285,** 627–643.
22. Woodward, C. (1993) Is the slow-exchange core the protein folding core? *Trends Biochem. Sci.* **18,** 359–360.
23. Li, R. and Woodward, C. (1999) The hydrogen exchange core and protein folding. *Protein Sci.* **8,** 1571–1591.
24. Clarke, J., Itzhaki, L. S., and Fersht, A. R. (1997) Hydrogen exchange at equilibrium: a short cut for analyzing protein-folding pathways? *Trends Biochem. Sci.* **22,** 284–287.
25. Englander, S. W. (1998) Native-state HX. *Trends Biochem. Sci.* **23,** 378.
26. Woodward, C. and Li, R. (1998) The slow-exchange core and protein folding. *Trends Biochem. Sci.* **23,** 379.
27. Clarke, J., Itzhaki, L. S., and Fersht, A. R. (1998) A reply to Englander and Woodward. *Trends Biochem. Sci.* **23,** 379–381.
28. Roder, H., Wagner, G., and Wuthrich, K. (1985) Amide proton exchange in proteins by EX1 kinetics: studies of the basic pancreatic trypsin inhibitor at variable p^2H and temperature. *Biochemistry* **24,** 7396–7407.

29. Pedersen, T. G., Thomsen, N. K., Andersen, K. V., Madsen, J. C., and Poulsen, F. M. (1993) Determination of the rate constants k_1 and k_2 of the Linderstrøm-lang model for protein amide hydrogen exchange: a study of the individual amides in hen egg-white lysozyme. *J. Mol. Biol.* **230,** 651–660.
30. Arrington, C. B. and Robertson, A. D. (1997) Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry* **36,** 8686–8691.
31. Arrington, C. B., Teesch, L. M., and Robertson, A. D. (1999) Defining protein ensembles with native-state NH exchange: kinetics of interconversion and co-operative units from combined NMR and MS analysis. *J. Mol. Biol.* **285,** 1265–1275.
32. Arrington, C. B. and Robertson, A. D. (2000) Kinetics and thermodynamics of conformational equilibria in native proteins by hydrogen exchange. *Methods Enzymol.* **323,** 104–124.
33. Arrington, C. B. and Robertson, A. D. (2000) Microsecond to minute dynamics revealed by EX1-type hydrogen exchange at nearly every backbone hydrogen bond in a native protein. *J. Mol. Biol.* **296,** 1311–1321.
34. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* **17,** 75–86.
35. Huyghes-Despointes, B. M. P., Scholtz, J. M., and Pace, C. N. (1999) Protein conformational stabilities can be determined from hydrogen exchange rates. *Nature Struct. Biol.* **6,** 910–912.
36. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* **269,** 192–197.
37. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of Rnase H. *Nat. Struct. Biol.* **3,** 782–787.
38. Jackson, S. E. (1998) How do small single-domain proteins fold? *Folding and Design* **3,** R81–R91.
39. Swint-Kruse, L. and Robertson, A. D. (1995) Hydrogen bonds and the pH dependence of ovomucoid third domain stability. *Biochemistry* **34,** 4724–4732.
40. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) Protein stability parameters measured by hydrogen exchange. *Proteins Struct. Funct. Genet.* **20,** 4–14.

41. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman and Co., New York, NY.
42. Davis, J. H. (1995) Refocusing revisited: an optimized, gradient-enhanced refocused HSQC and its applications in 2D and 3D NMR and in deuterium exchange experiments. *J. Biomol. NMR* **5,** 433–437.
43. Hwang, T. L., van Zijl, P. C. M., and Mori, S. (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical Exchange (CLEANEX-PM) approach with a fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11,** 221–226.
44. Baxter, N. J. and Williamson, M. P. (1997) Temperature dependence of ^{1}H chemical shifts in proteins. *J. Biomol. NMR* **9,** 359–369.
45. Forsyth, W. R. and Robertson, A. D. (1996) Intramolecular electrostatic interactions accelerate hydrogen exchange in diketopiperazine relative to 2-piperidone. *J. Am. Chem. Soc.* **118,** 2694–2698.
46. Evans, J. N. S. (1995) *Biomolecular NMR spectroscopy*, Oxford University Press, Oxford, U.K.
47. Westcott, C. C. (1978) Electrodes, in *pH Measurements*, 1st ed., Academic Press, Inc., Orlando, FL, pp. 41–78.
48. Bates, R. (1973) pH standards, in *Determination of pH: Theory and Practice*, 2nd Ed., John Wiley and Sons, New York, NY, pp. 59–104.

10

Molecular Dynamics Simulations of Protein Unfolding/Folding

Valerie Daggett

1. Introduction

Major developments in the last decade have made it possible to study entire folding pathways at the level of individual residues and atoms. This is largely the result of developments in two areas that bring together complementary experimental and theoretical approaches. The first is the use of recombinant DNA methods to produce small proteins with desirable characteristics for the study of folding. This technology also allows the production of mutants to probe the folding process and examine the contributions of individual residues to the kinetics and thermodynamics of folding. The second is the development of molecular dynamics methods and the parallel advances in computing that make it possible to simulate protein denaturation at atomic resolution.

Besides providing an unprecedented view of conformational transitions during protein folding, these developments allow an in-depth characterization of protein folding/unfolding transition states. For full elucidation of the protein-folding process, all states along the pathway must be characterized, including the transient and somewhat nebulous transition state. Until recently, folding studies have

From: *Methods in Molecular Biology, vol. 168: Protein Structure, Stability, and Folding*
Edited by: K. P. Murphy © Humana Press Inc., Totowa, NJ

focused primarily on intermediate states to piece together pathways. However, since the demonstration that proteins can fold in a two-state manner without the formation of stable intermediates ***(1)***, more attention has been directed to the only "state" between the native (N) and denatured (D) states. The protein engineering method (Φ-value analysis) is the only experimental technique currently available for probing the transition state in depth, and by combining it with molecular dynamics simulations, we can obtain an atomic-level view of these transient protein folding/unfolding states.

In this way, experiment and theory can be combined to produce a description of protein folding from the native to the denatured states of small proteins at atomic resolution. NMR and Φ-value analysis provide the average structures and properties of the ensembles of molecules that constitute the ground, intermediate, and transition states. Molecular dynamics simulations provide the structures of individual molecules over a continuous reaction coordinate. There is sufficient agreement between theory and experiment where they meet at discrete conformational states—such as intermediate and transition states—that the simulated pathways seem reliable. Thus, experiment takes the lead with information about discrete states, thereby revealing glimpses into the folding process, while molecular dynamics simulations provide high-resolution structural and dynamic information for these states as well as the links between them. This chapter focuses on the characterization of protein folding transition states and unfolding/folding pathways by molecular dynamics simulations and describes how the results are validated through experiment.

2. Characterization of the Protein Folding Process

The full characterization of the protein folding process has been complicated by a variety of factors. One is that the free-energy difference between the native, denatured, and any other states that may be populated at equilibrium is small (~1–15 kcal/mol). As such, slight changes in conditions can have profound effects on these states, their interconversions, and the ability to characterize them experimentally. Furthermore, there is no guarantee that a particular state is relevant to the folding process, even when conditions favor this state over others. Moreover, as the protein becomes less ordered, moving away from the native state, the concrete structural information that can be obtained becomes severely limited. Transition states are only transiently populated and cannot be studied by conventional structural techniques. Instead, transition states can only be studied through kinetics, which also provide the only way to ensure that any state is relevant to the folding process.

While we would like to describe the folding reaction coordinate in terms of discrete structures becoming progressively more ordered during folding, we are not actually dealing with discrete entities. Instead, we have "states" or ensembles comprised of many different conformations with similar properties. This is analogous to thermodynamic states being comprised of many quantum states. Experimentally, then, the average properties of this huge ensemble are monitored. Given the many weak noncovalent interactions that define and stabilize proteins, even thermal energy can be sufficient for conformational interconversions within and between these states. It is desirable—yet problematic—to make the leap to the molecular level and discuss these states in terms of specific structure. To this end, molecular dynamics simulations of unfolding can be performed to complement the experimental information, and these provide a structural framework for interpretation of the data.

2.1. Molecular Dynamics Simulations of Protein Unfolding

The conformational properties of macromolecules cannot be studied using the types of analytical methods suitable for small molecules. Instead, computer simulation methods are used—particularly molecular dynamics, which by providing atomic positions as a function of time can provide a realistic and complete picture of protein motion. The all-atom description of a protein immersed in a solvent can provide a detailed picture of the evolution of the system with pathways dictated by the forces responsible for protein folding and stability. All interactions—both native and nonnative—and collective ensembles of conformations can be generated. Therefore, molecular dynamics simulations represent another approach that can both complement and extend the information about the folding pathway, so that links between the discrete conformational states probed experimentally may be elucidated. All of this is predicated on the ability of the simulations to reproduce the properties of experimentally targeted ensembles.

The technique of molecular dynamics requires a well-defined starting structure and a potential energy function with parameters derived to reproduce structures and energy trends in various model systems. During the simulation, the atoms move as a result of their own kinetic energy and the forces exerted upon them by all other atoms. One can then monitor specific interactions both geometrically and energetically to investigate structural transitions and the mechanism by which they occur. The results of a molecular dynamics trajectory can be viewed as a movie. This process reveals tremendous detail about the conformational transitions that a molecule undergoes, the interactions that stabilize a structure, and the influence of solvent on the dynamics. Given the absolute importance of solvent in determining the conformational properties of proteins, it is imperative to include solvent in the simulation to realistically model denaturation (*see* **Fig. 1**).

This general approach has been used to study the unfolding process and characterization of predominantly early unfolding events, including the properties of transition states ***(2–10)*** and intermediate states ***(11–18)***, the details of the unfolding process (***11***, ***19–26*** and the refs. cited above), the energetics of folding ***(27–30)***, and more recently the denatured state ***(6,10,31–32)***. Most unfolding simulations have employed high temperature to destabilize the native state, but simulations including chemical denaturants are becoming more common ***(14,24,26)***. It is also possible to take structures from unfolding simulations, return them to folding conditions, and observe folding events such as hydrophobic collapse and secondary

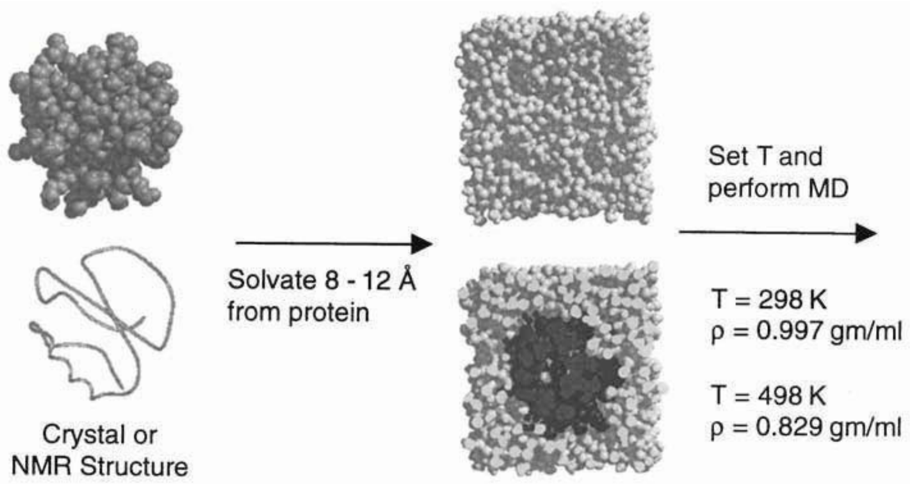


Fig. 1. Setting up a high-temperature molecular dynamics simulation, using CI2 as an example.

structure formation ***(14,33,34)***. Unfortunately, the μs simulation of Duan and Kollman ***(34)***—currently the longest simulation performed to date by two orders of magnitude—failed to observe folding events not accessible on the ns time-scale ***(14,33)***. Because of these severe sampling limitations under folding conditions, we must focus on unfolding events.

There are numerous advantages to studying unfolding as opposed to folding. Simulations begin from a well-defined starting structure—a crystal or nuclear magnetic resonance (NMR) structure—which improves the odds of sampling relevant regions of conformational space. Another advantage to studying unfolding is that the full reaction coordinate from the native to denatured states can be explored. The temperatures employed to unfold the protein are typically very high (498–600 K or 225–327°C). It is important to note that if one is careful to set the solvent density to the experimental value of the temperature of interest, then the excess pressure can be reduced and the water will remain in liquid form up to 498 K. While these conditions are severe, a 200° increase in temperature from 298 → 498 K corresponds to a ~30% increase in the atomic velocities. Recent studies also demonstrate that the pathway of unfolding is temperature-independent—temperature increases merely accelerate the process ***(35,36)***. Because of the improved sampling under denaturing conditions and the diminished probability of becoming trapped, multiple unfolding simulations can be performed in which relevant regions of conformational space are sampled—a process that provides more information than a single longer "folding" simulation. In monitoring unfolding, several trajectories are generally sufficient to infer information about a protein system, although one must take care in generalizing results from one or a handful of simulations. The unfolding process is of interest because of the connections with protein degradation and disease and the principle of microscopic reversibility dictates that the pathways of folding and unfolding must be the same under the same conditions. Therefore, the mechanism of folding can be explored from both directions, and information obtained from the study of unfolding can be used to understand the mechanism of folding.

As with experiment, the study of transition states presents problems for simulation studies. The transition state for folding/unfolding is an ensemble of high free-energy structures. Unfortunately, even if a reasonable unfolding pathway can be simulated with molecular dynamics, the calculation of free energies for such a complicated process is not possible. So, we rely on structural properties—the strength of force-field methods—to identify the transition state in a simulation. Using this approach, structural attributes of the transition-state ensemble can be precisely delineated, yet there is no guarantee that the ensemble identified is the state of highest free energy. Since the transition state is kinetically and thermodynamically unstable, we expect the structure of the protein to change rapidly once it passes the major transition state. Therefore, we define the major transition state in the simulations as the ensemble of structures populated immediately before the onset of a large structural change. The easiest way to find the transition-state regions is through a conformational cluster analysis ***(2,3)*** as discussed further in **Subheading 3.1.**

for chymotrypsin inhibitor 2 (CI2). Because we cannot determine free energies in an unrestrained MD simulation, the current definition of the transition state—as written—is not precise or rigorous, thus making extensive comparison with experiment essential. Also, because of sampling limitations with molecular dynamics individual simulations are not expected to capture all of the features of the true conformational ensemble examined experimentally.

2.2. Protein Folding/Unfolding Transition States

The transition state of protein folding/unfolding differs from an elementary chemical reaction in many ways. First, low-energy nonbonded interactions are broken in the transition state, instead of the covalent bonds characteristic of a chemical reaction. Secondly, there may be many energy barriers along a single unfolding pathway. If one barrier has a much higher activation energy than the others, this reaction step will be rate-limiting and, by convention, is considered the transition state of the overall reaction ***(37)***. Thirdly, chain conformational entropy is the main driving force for unfolding. The delicate balance of enthalpy and entropy makes the folded state of a protein stable under one set of conditions and unstable under another. In addition, the importance of entropy means that the major transition state is the ensemble of structures with the highest free energy.

Another consideration is that there are many possible pathways from the native to the unfolded state. Some of these pathways may be energetically similar and of equal importance; thus, the structure of the transition state may not be unique. Even if only a single pathway is accessible under certain conditions and one can identify the transition state for the overall process, the transition state will be an ensemble of structures. This is analogous to a simple chemical reaction that has a relatively flat and broad saddle-point region, both in the direction along the reaction coordinate and the direction perpendicular to it. The transition state in protein unfolding potentially

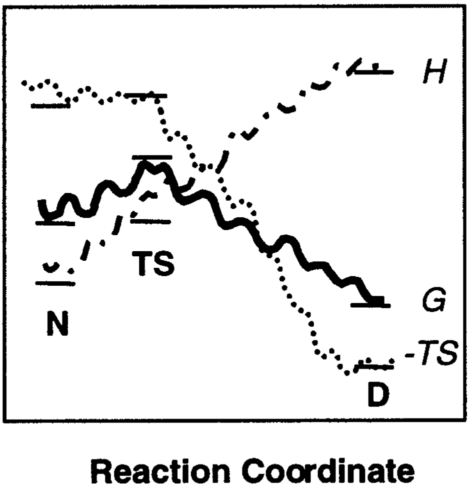


Fig. 2. Schematic representation of how increases in enthalpy during unfolding in the absence of large changes in entropy can lead to a free-energy maximum, and thus transition state.

involves the entire protein and, in any case, will not be localized to a particular bond as in a chemical reaction.

Nearly all of the protein transition states studied to date are relatively structured and native-like, but with disrupted packing. Given such a scenario, the free-energy maximum of the transition state can be explained by an increase in enthalpy combined with only a minor increase in entropy relative to the native state. Disruptions to the packing of the molecule in the transition state lead to increases in enthalpy. If these disruptions do not lead to severe differences in the structure, particularly the main chain, then the increase in entropy should be minor. In addition, a modest increase in protein entropy at this stage is countered by a drop in water entropy caused by exposure of some nonpolar groups. This situation should lead to a free-energy barrier. Upon further unfolding, main-chain and side-chain entropy will increase, thereby lowering the free energy again. Such a trade-off of effects is depicted schematically in **Fig. 2**.

2.2.1. Φ-Value Analysis for Detailed Characterization of Transition States

Protein engineering techniques have had a profound effect on the protein-folding field. For example, one can now easily introduce spectroscopic probes into a protein, thus providing the means to monitor the folding process when it may not have been possible previously. However, a more comprehensive mapping of the folding process can be obtained by systematic study of the energetic consequences of introducing mutations throughout the protein—termed the protein engineering method or Φ-value analysis. Since one evaluates the effects of mutations by both kinetic and equilibrium experiments, the method yields valuable information about transitory states relevant to the folding/unfolding process, such as transition and intermediate states. The mutations act as probes, so that the structure at the site of mutation can be inferred from the energetics. More specifically, characterization of a structure of interest is based on a ratio of the destabilization energies upon mutation: $\Phi = \Delta\Delta G_{TS-D} / \Delta\Delta G_{N-D}$ (*see* **ref. *38*,** and **Fig. 3**). Consider a case where, in the transition state of unfolding, the structure of the protein at the site of mutation is the same as in the native state. Then, the protein is immune to the effect of the mutation until after the major transition state, and the transition state is destabilized by exactly the same amount as the native state—that is, $\Delta\Delta G_{TS-D} = \Delta\Delta G_{N-D}$ and $\Phi = 1$ (this is depicted schematically in **Fig. 3**). Conversely, a Φ value of 0 means that the structure of the transition state at the site of mutation is the same as in the denatured state (*see* **Fig. 3**). Intermediate values represent structures that are partially unfolded in the transition state (or a mixture of folded and unfolded).

There are a number of assumptions associated with this approach. The first is that structure is inferred from energetics. While we know that structure and energy must be linked, they are not linearly related. The precise mathematical relationship between the two is unknown—which is mostly a problem in interpreting partial Φ val-

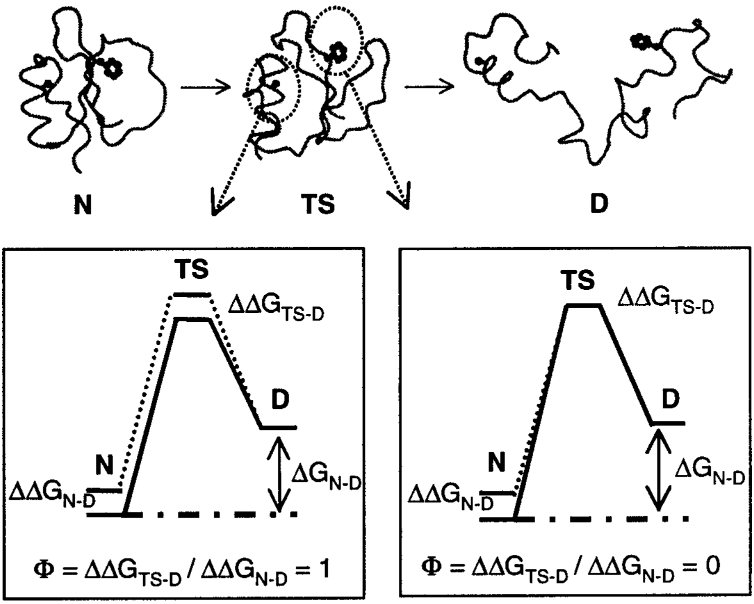


Fig. 3. Schematic representation of the protein engineering approach and how structure is inferred from energetics.

ues. Another assumption is that the mutation merely examines what occurs in the wild-type transition state (or intermediate). As such, the mutation should not change the pathway of folding/unfolding or introduce new interactions. Therefore, the choice of mutations is critical. In general, nondisruptive, conservative changes are desired. Nonetheless, the fundamental approach appears to be sound based on the consistency of the results of multiple mutations both at a site and around a site.

The protein engineering method and Φ-value analysis of transition states is gaining popularity, and has now been applied to several systems: barnase ***(37–42)***, chymotrypsin inhibitor 2 (CI2) ***(39,43–46)***, barstar ***(47)***, CheY ***(48)***, FKBP12 ***(8)***, the activation domain of procarboxypeptidase A2 ***(49)***, the SH3 domains of α-spectrin ***(50–51)***, and Src ***(52)***, λ repressor ***(53)***, protein L ***(54)***, and p13suc1 ***(55)***. These studies suggest that the transition state is native-like, and they also provide site-specific information regarding the integrity of interactions in the transition state. However, although Φ-value analysis provides extensive information about the transition state, combining it with molecular dynamics simulations can take it a step further to obtain structures to go with the energetics. Simulations can supplement the information obtainable from experiment and address the validity of the assumptions.

2.3. The Denatured State: The End Point of Unfolding

A complete description of the folding process cannot stop with the transition state, even in a two-state system. Characterization of the denatured state is also required. Advances in NMR have made it possible to obtain unprecedented information about the denatured states of proteins (*see* reviews ***56–58*** and the original studies ***59–79***). These studies have demonstrated that the denatured state is best described as an ensemble of conformations interconverting much faster than 10^3 s^{-1}. The structures that comprise this ensemble do not usually represent all—or a random sampling—of conformational space. Instead, they display structural preferences referred to as residual structure. The extent and nature of the residual structure varies with the protein and the environmental conditions, and typically involves the formation of transient hydrophobic clusters and/or residual but unstable secondary structure and turn structures.

While these experimental methods have revolutionized our view of the denatured state, the data cannot be transformed into molecular models, and they generally only yield limited amounts of information about the actual molecular details of the motion experienced by a protein. Therefore, as with transition states, theoretical studies are needed to better elucidate the properties of the denatured ensemble.

3. Protein Folding at Atomic Resolution: Combining Φ-Value Analysis, Molecular Dynamics Simulations, and NMR

From the preceding discussion, the complementary and even synergistic nature of experimental and simulation studies should be evident. Experimental reports describe the average properties of a large ensemble (approx 10^{17}–10^{19} molecules), typically over time. Molecular dynamics simulations, on the other hand, provide detailed information for a single molecule, a component of the ensemble. A combination of these approaches can provide a more complete picture of folding, and can also take advantage of the strengths of the different methods. For example, experiment provides averages that may or may not adequately reflect the true properties of an individual protein, despite the fact that such results are always discussed in these terms. Also, the experiments cannot provide molecular structures. Molecular dynamics simulations can provide such information but the simulations are meaningless without the link to experiment. Thus, the interrelated nature of simulation, in assessing the assumptions of the protein engineering method and providing structures to augment the energetic description, and experiment, and in judging whether the simulations are reasonable, provides more confidence in the resulting information about folding. This combination yields tested and testable molecular models of states that evade characterization by conventional methods. This chapter explores the combined use of these methods to map folding/unfolding pathways at atomic resolution using chymotrypsin inhibitor 2 (CI2) as an example. This protein represents the most thoroughly studied system for folding from both the experimental and theoretical points of view.

3.1. Chymotrypsin Inhibitor 2: A Test Case

CI2 contains 64 residues, of which there is a single α-helix and a mixed parallel and antiparallel β-sheet (**Fig. 4**). CI2 was the first protein shown to fold in a two-state manner ***(1)***, and it is a good model for studying elementary folding events. As a two-state folding protein, characterization of the transition state is crucial, since it represents the only observable sp between the native and denatured end points:

$$\mathrm{D} \xleftrightarrow{\mathrm{TS}} \mathrm{N}$$

Also, because of its lack of subdomains and relatively uniform structure—with essentially the entire chain contributing to a single mod-

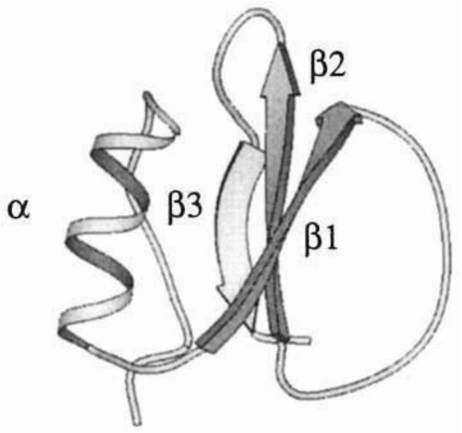
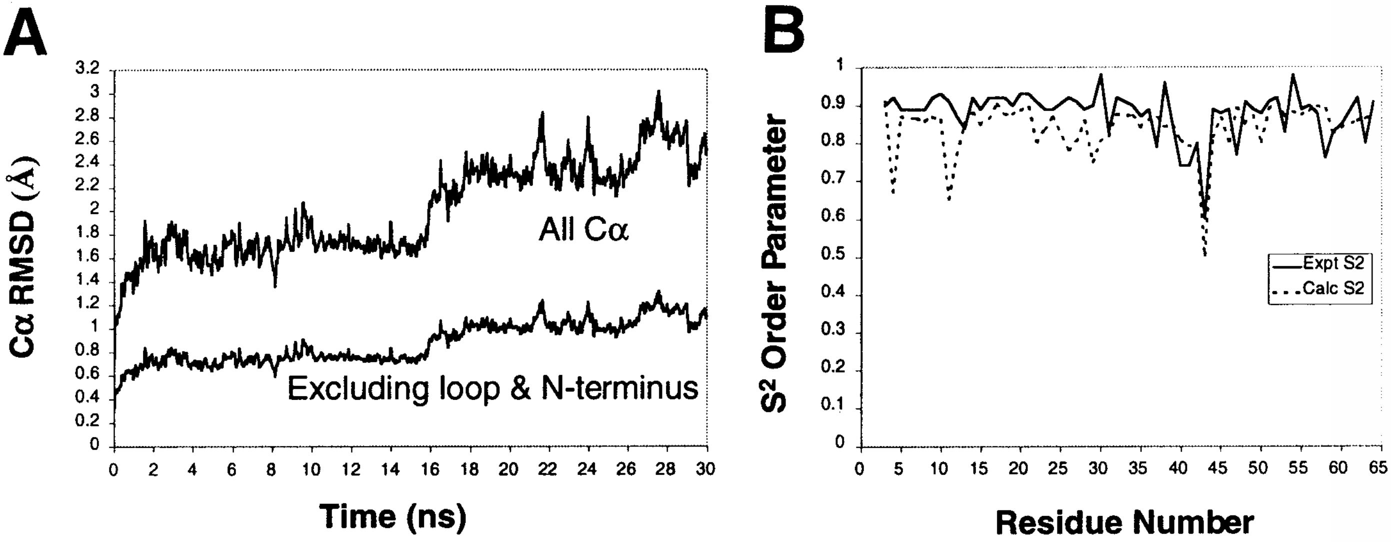


Fig. 4. The main-chain fold of CI2 with secondary structure labeled.

ule of structure—CI2 represents a basic folding unit, or foldon ***(80)***. Our task is to characterize the native, transition, and denatured states by both experiment and simulation, and to use the simulations to connect these states for a full description of the folding pathway.

3.1.1. The Native State of CI2

An important and necessary first step in simulating protein unfolding pathways is a demonstration that simulation of a protein under native conditions—using the same force field and protocols as will be used for denaturation—reproduces the native fold, dynamic behavior, and other properties that define the native state. In order to test whether one can perform long and restraint-free MD simulations from either the crystal or NMR structure and reach an equilibrium structure that would satisfy the observed nuclear Overhauser effect cross-peaks (NOEs) indicative of close H . . . H distances, a long native-state simulation of CI2 was performed. The solvated protein was simulated for 5.3 ns (5.3×10^6 time-steps) at 298 K in water ***(81)***. The backbone stayed near both the crystal and NMR structures (within ~1.5 Å Cα root mean square deviation [RMSD]). In fact, a more recent 35-ns native simulation of CI2 yields similar results; the Cα RMSD remains < 2 Å, and the RMSD


A
Cα RMSD (Å)
3.2
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Time (ns)
All Cα
Excluding loop & N-terminus
B
S² Order Parameter
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65
Residue Number
Expt S2
Calc S2

of the core (excluding the active-site loop and N-terminus) is ~0.8-1 Å (*see* **Fig. 5**) ***(36)***. The MD simulation reproduces the majority of the experimentally measured coupling constants, short- and long-range NOEs, and the amide hydrogen-exchange behavior. Interestingly, the loop that binds in the protease active site is quite mobile, and this heightened mobility is in agreement with the results of NMR relaxation experiments (***81–82***, and **Fig. 5**).

3.1.2. The Transition State of CI2

The folding ↔ unfolding of CI2 fits a two-state model both kinetically and thermodynamically ***(1)***. The structure of the transition state of folding/unfolding has been studied with a variety of techniques, including a Φ-value analysis using >100 mutations spanning the length of the protein ***(43–45)***. In addition, both unfolding and refolding were investigated under a variety of conditions, and the results are independent of direction, indicating that the transition states for folding and unfolding are equivalent.

There is an increase of approx 40% in the solvent-accessible surface area of the transition state of CI2 relative to its native state, as judged by β_T ***(43)***. Roughly 50% of the hydrophobic residues are buried in the transition state—$\Delta C_{P(TS-D)}/\Delta C_{P(N-D)} = 0.5$ ***(83)***. A plot of the change in activation free energy of unfolding ($\Delta\Delta G_{TS-N}$) vs the change in the folding free energy ($\Delta\Delta G_{N-D}$) upon mutation fits well to a Brønsted plot with a slope of 0.7 ***(39,46)***, suggesting that ~30% of the extent of side-chain interactions are intact in the transition state. The folding of CI2 is highly cooperative by a variety of measures, demonstrating the tight link between secondary and tertiary structure throughout the protein. For example, hydrogen

---

Fig. 5. *(opposite page)* **(A)** The Cα root-mean-square deviation from the crystal structure of CI2 as a function of simulation time for all residues and excluding the active site loop and N-terminus. **(B)** The calculated and experimental (Shaw et al., 1995) S^2 order parameters reflecting main-chain mobility. Mobility of the N-H bond vectors increase as the S^2 values decrease.

exchange in CI2, as probed by increases in denaturant and temperature, indicates that exchange behavior is the same for all of the hydrogens and that they merge with the global unfolding reaction at high denaturant concentration ***(83)***. Real-time NMR folding experiments show that all of the peaks throughout the protein appear at identical rates ***(84)***. Furthermore, these rates mirror that of global folding as monitored by stopped-flow Trp fluorescence.

These overall features of the transition state can be supplemented by the more site-specific information that can be obtained from a Φ-value analysis. The Φ-values for CI2 tend to fall between 0.2 and 0.5 ***(43–45)***. Some values are higher; these are for residues in the α-helix—particularly at the N-terminus, and residues in the β-sheet that dock to the helix.

In an attempt to obtain molecular models for the transition state of folding/unfolding of CI2, four unfolding simulations of CI2 (beginning from the crystal structure and various NMR structures and termed MD1–MD4) were performed, and a transition state was identified from each (*see* **Fig. 6**). The simulations were done in parallel with the experimental studies in a blind manner—performed as predictions rather than fits to experiment. The transition states identified in the four independent simulations (TS1–TS4) are similar overall, and the unfolding pathways only diverge past the transition state (*see* **Fig. 6**). The transition states have the following characteristics: the hydrophobic core is considerably weakened; the secondary structure—particularly the β-sheet—is frayed; and packing of the secondary structure is highly disrupted. (*see* **Fig. 7**). The overall structure of the transition state is closer to the native than the unfolded state.

For comparison with the experimental Φ values, "computer-mutations" were made to the transition-state structures, and the difference in packing contacts between the wild-type and mutant proteins in the transition and native states were evaluated to determine a Φ_{MD} value ***(2)***. This approach provides good agreement with experiment for the 11 hydrophobic deletion mutants described by Jackson and colleagues ***(43)***; the correlation coefficient for comparison of Φ_{MD} and the experimental Φ-values is 0.88 for the most thoroughly studied transition-state model, TS1. In fact, the best

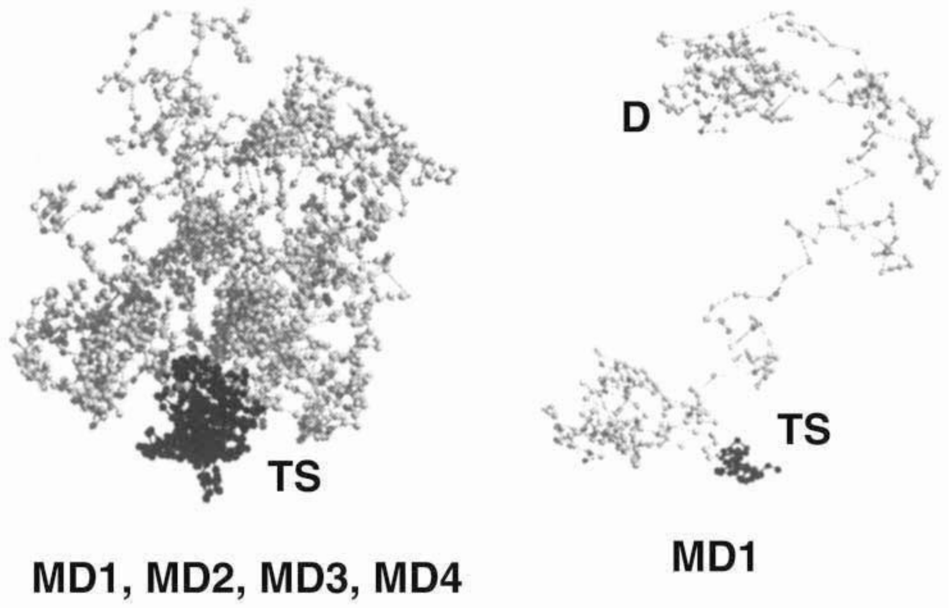


Fig. 6. The conformational clustering projections used to identify possible transition-state regions for CI2. On the left is an overlay of the first ns of four simulations showing how divergence in pathways occurs after the major transition state. On the right is a projection for a single simulation, MD1. The clustering projection is a reduction of the dimensionality of the RMS deviations between all structures in a simulation with all other structures (i.e., 15,000 × 15,000 for a 3 ns simulation). The distance between the points is approximately equal to the actual RMS deviation between the structures. Similar structures, regardless of time, form clusters.

agreement with experiment is obtained when the four transition states are pooled ($R = 0.94$), highlighting the fact that the transition state is an ensemble of related structures ***(3)***. In more recent studies, we evaluate only the structure of the wild-type transition-state structures, which allows us to test the assumption that the mutations are merely probes of what occurs in wild-type protein. The structure of the wild-type transition state can be evaluated semi-quantitatively using local structure indices, termed *S* values. For each residue, an *S* value is calculated that is a product of the local percentage of tertiary and secondary structure in the transition-state structures relative to the native state ***(4)***. The calculated *S* values agree well

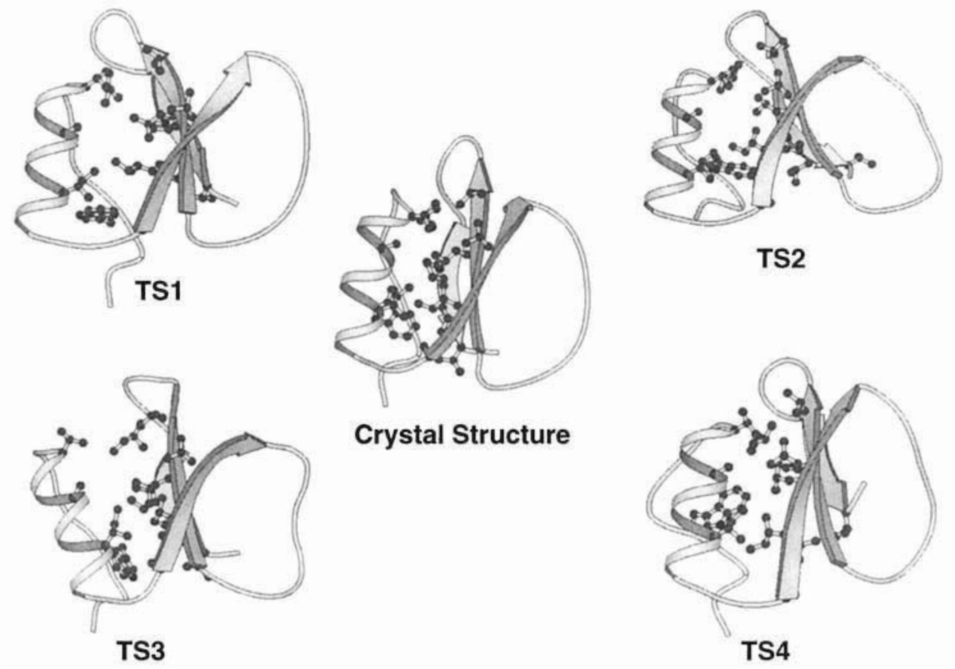


Fig. 7. Secondary structure and packing interactions in the transition-state models of CI2.

with the experimental values (*see* **Fig. 8**) across the protein, giving a correlation coefficient of 0.9. The agreement lends support to the assumption that the protein engineering approach does not dramatically change the folding process, and can report on the behavior of the wild-type protein. Also, we note that independent unfolding simulations of CI2 by Lazaridis and Karplus ***(7)*** using a different force field are consistent with the results described here.

Despite the good agreement between simulation and experiment for the transition state of folding/unfolding of CI2, the high temperature used in the simulations and the limited sampling are cause for concern. For example, Oliveberg et al. ***(85)*** have shown that the transition state of CI2 becomes more native-like at high temperature, demonstrating the value of further experimental tests of the simulations. To do so, the simulated structures of the transition state were used to design faster folding mutants of CI2. The models pin-

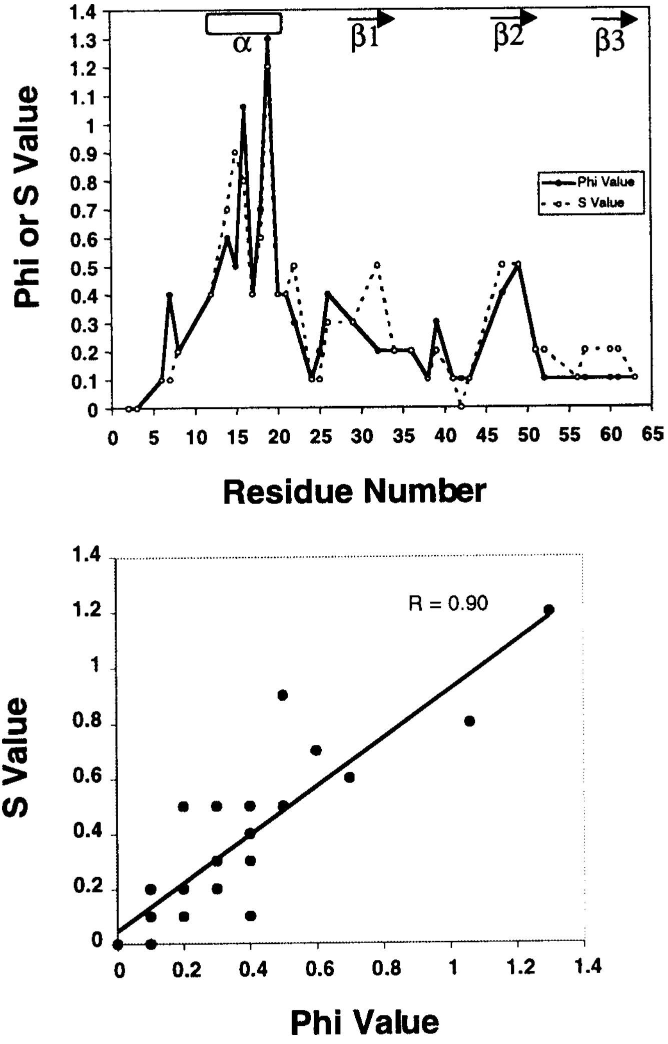
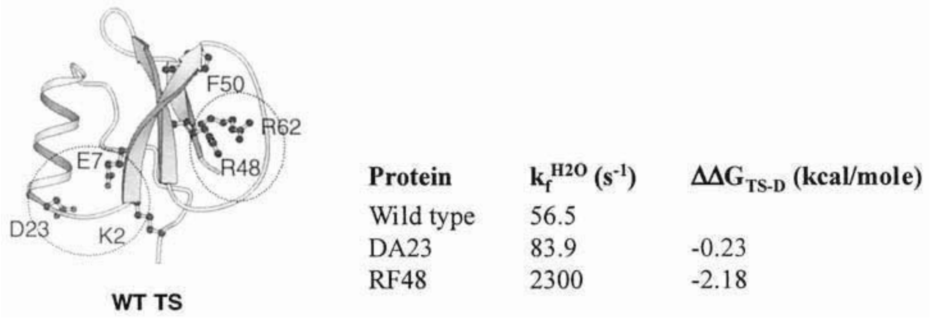


Fig. 8. Comparison of experimental Φ-values and the average calculated S values for the CI2 transition state. The average is taken over the TS1-TS4 ensembles.

point a number of unfavorable local interactions at the carboxyl-terminus of the α-helix and in the protease-binding loop region of CI2. Thus, the prediction is that if unfavorable interactions are removed via mutation, folding will become more rapid.



Protein	k_f^{H2O} (s^{-1})	$\Delta\Delta G_{TS-D}$ (kcal/mole)
Wild type	56.5	
DA23	83.9	-0.23
RF48	2300	-2.18

Fig. 9. Unfavorable electrostatic interactions in the transition state of CI2 (a negative charge at the C-terminus of the helix and charge repulsion in the active-site loop).

The first region investigated was the C-terminus of the helix. Asp 23 stabilizes the native protein by creating a salt bridge with Lys 2. Yet the presence of an aspartate at this position in an isolated helix is destabilizing through unfavorable interactions with the carbonyl groups at the end of the helix—the helix macrodipole. The simulated transition state effectively has an isolated helix when this salt bridge is broken (*see* **Fig. 9**). A DA23 mutation is predicted to fold faster than wild-type CI2 through stabilization of the transition state for folding. This is demonstrated as the refolding rate constant increases from 56 for wild-type, to 84 s^{-1} for DA23 (***86)*** (*see* **Fig. 9**). These increases are especially significant considering that overall destabilization of CI2 generally leads to a decrease in the rate constant for folding ***(87)***.

The second region (the minicore) investigated contains a cluster of positive charges, comprised of Arg 43, Arg 46, Arg 48, and Arg 62 (*see* **Fig. 9**). The hydrophobic side chains of these residues stack, leading to their guanidinium groups being close to one other and causing electrostatic strain. This strain is partly relieved by a network of hydrogen bonds with the carboxylate of the C-terminal residue, Gly 64, in the native state. This loop region is expanded and more loosely packed in the transition state (*see* **Figs. 7** and **9**). The transition-state models show three or four of the Arg residues in proximity, and the native salt bridges and favorable ionic interactions are not well-formed. The removal of some of the unfavorable electrostatic interactions between the positively charged guanidinium groups, and improvement of nonpolar packing in the region, would therefore be expected to stabilize the transition state. An RF48 mutation was made, and the rate of folding increased from 56 to 2,300 s^{-1} to yield the fastest folding form of CI2 thus far.

The mutations described here, and others (*see* **ref. *86***) were designed to yield faster folding versions of CI2 based solely on the MD-generated transition-state models. Removal of unfavorable interactions identified in the models increased the rate of folding. This completes the cycle by first performing the simulations and experiments separately, by comparing the results and showing the complementary nature of the two approaches, and finally by making predictions based on the model structures and testing those predictions in the lab.

3.1.3. The Denatured State of CI2

The denatured state of CI2 is largely unstructured, as probed by NMR studies of the full-length protein under different conditions and various fragments ***(88–91)***. However, there appears to be some tendency for very weak native helical structure and some weak clustering of hydrophobic residues, particularly near the center of the protein—as revealed by deviations in NMR chemical shifts from random coil values. The $^3J_{NH\text{-}C\alpha\, H}$ coupling constants are 7.2, on average, and relatively constant along the sequence, suggesting that there is no residual structure or that the residual structure is weak enough to be averaged out. Also, the pK_a values for the acidic residues are weakly perturbed from their values in model compounds—0.3 lower—suggesting that there are some weak electrostatic interactions in the denatured state ***(92)***. For comparison, the pK_as of these residues are ~2 units lower in the native state. Thus, the available experimental evidence suggests that there may be some very weak residual structure in the denatured state of CI2, but it is largely unstructured and expanded.

The simulations described here, addressing the transition state, were carried out until the protein unfolded to investigate the nature

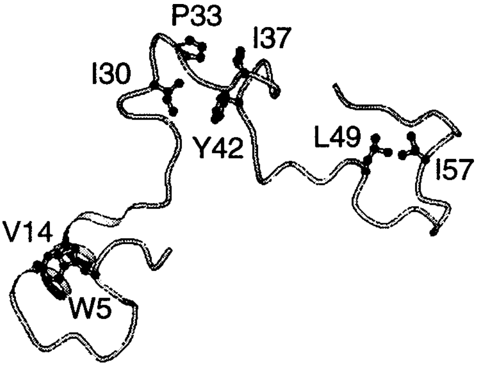


Fig. 10. The denatured state of CI2 is very expanded, with some dynamic hydrophobic clusters and some residual native helical structure. The structure is a 3-ns snapshot from the MD1 simulation of CI2.

of the denatured state. The simulated denatured state is expanded with little persistent secondary structure and few tertiary contacts (*see* **Fig. 10**). There is some dynamic, residual native helical structure, but the β-sheet is totally destroyed in all simulations. There are some dynamic hydrophobic clusters in the denatured ensemble, and the more persistent ones are found in the center of the protein (*see* **Fig. 10**).

The $^3J_{NH\text{-}C\alpha H}$ coupling constants were calculated from the simulations for the denatured state ensemble from the average ϕ dihedral angles ***(93)***. The range of values (6–10 Hz) is the same for both the simulations and the NMR data, with average values of 7.4 and 7.2 Hz, respectively. Simulation and experiment failed to show a decrease in the coupling constants in the areas of residual structure. These values demonstrate how an experimental observable may not reveal residual structure when averaged over an entire ensemble.

Besides the secondary structure, the experimental studies also suggest that there is hydrophobic clustering between residues Ile 29-Ile 37, based on the high number of NOEs and various chemical shift values ***(88,91)***. This region of CI2 is highly hydrophobic (the sequence is IIVLPVGTI). In the native protein, this region encompasses the C-terminal half of β-strand 1 and part of the active-site loop. In the simulations, this region makes a number of local hydrophobic contacts as well as with other more distant hydrophobic residues in the molecule. A representative denatured structure of CI2 is displayed in **Fig. 10** with the hydrophobic interactions in the 29–37 region highlighted. The hydrophobic interactions within this region facilitate formation of the turn coming from β-strand 1 into the active-site loop, which triggers formation of the active-site loop and hydrophobic minicore.

3.1.4. Summary of the Folding Pathway of Chymotrypsin Inhibitor 2

Both experimental and theoretical studies have been undertaken in an attempt to characterize the transition state of protein folding/unfolding of chymotrypsin inhibitor 2. The two methods generally produce quite similar results. Both methods indicate that there is partial structure formation in the transition state: the Φ values are intermediate between values indicative of fully formed or fully disrupted structure. The α-helix is the most structured region of the protein, although it is very mobile. The β-sheet is substantially disrupted, with essentially complete loss of structure at the periphery away from the core (**Fig. 6**). Furthermore, both the experiments and simulations suggest that there is a single family of transition states as opposed to parallel pathways and parallel transition states.

It has been proposed that Ala 16 is a nucleation site in folding ***(87,94)***. The simulations provide some clues as to what may be occurring in the vicinity of this residue during the folding process. Ala 16 retains some native interactions with neighboring core residues, most notably with Leu 49 and Ile 57, although packing contacts are diminished relative to the native state. While these interactions may indicate that the α-helix is docked onto the β-sheet in the transition state, the transition-state models indicate that the docking involves only a portion of the β-sheet. The Φ value analysis supports this idea, as only Ala 16, Val 19, and Leu 49 have Φ values ≥0.5 (**Fig. 8**). In the case of Val 19 the high value is a result of very local packing effects on the surface, which are unrelated to folding per se ***(3)***. Interestingly, there is strain between Ala 16, Leu 49, and Ile 57 in the folding nucleus of CI2, which is relieved during the process from the native to the transition state ***(86)***.

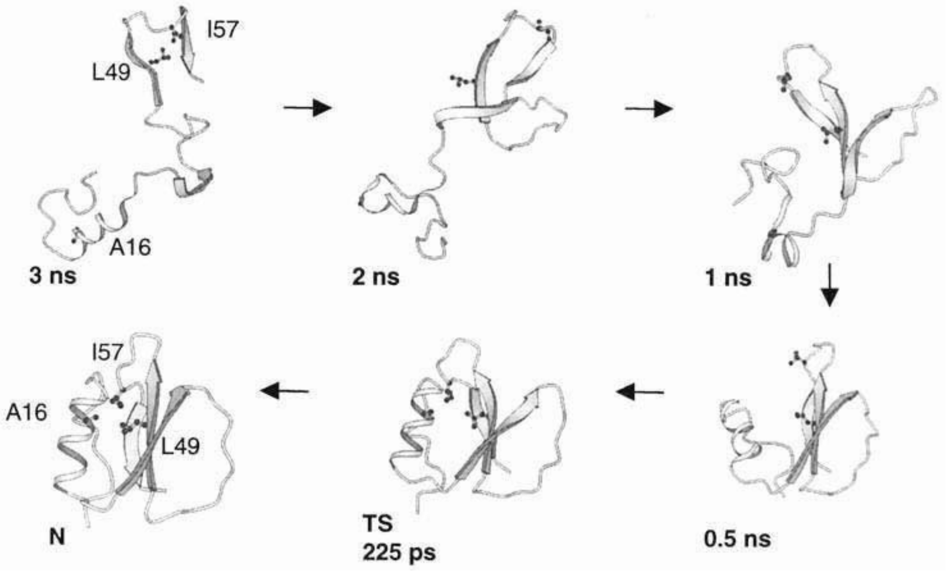


Fig. 11. Residues important in the transition state (A16, L49, and I57) only come together in the transition state.

The overall picture from the experimental and theoretical studies is that CI2 folds by a nucleation-condensation/collapse mechanism. As suggested by the Φ-value analysis and demonstrated directly in the simulations, expansion and loss of secondary structure occur concomitantly ***(3)***. Particular patches on the α-helix and the β-sheet are important in nucleating the core, and structure is looser radiating away from this area (**Fig. 7**). Subsequent events in folding involve condensation about this region, which in turn leads to consolidation of secondary structure. These patches of structure do not exist in the denatured state (*see* **Figs. 10** and **11**) and fragment studies show that there is little drive for local structure in the denatured state ***(87,89)***. Instead, the nucleation site remains embryonic until sufficient long-range contacts are made. These patches of structure only come together in or near the transition state (*see* **Fig. 11**). Given the nonlocal nature of these tertiary contacts, they are necessarily linked to secondary structure for productive folding to occur. After coming together, the collapse and condensation of structure around these patches are rapid.

4. Conclusions

It is now possible to use molecular dynamics simulations to elucidate protein-unfolding pathways at atomic resolution. The resulting pathways and associated conformational ensembles are in both qualitative and quantitative agreement with experiments that examine both unfolding and refolding. Furthermore, the simulations can be used to design faster folding proteins. The goal of these simulations is to sample experimental relevant regions of conformational space—a reasonable goal, since real proteins do not sample all of conformational space. However, while the simulations can provide exquisite detail for dominant components of the conformational ensembles explored experimentally, they severely undersample conformational space relative to experiment. In any case, the combination of Φ-value analysis, NMR, and molecular dynamics simulations is highly synergistic and provides an unprecedented view of protein folding and unfolding.

Acknowledgments

Support for this work was provided by the NIH (GM 50789). UCSF MidasPlus (96) was used to make **Figs. 1** and **6**, and Molscript ***(97)*** was used for **Figs. 4**, **7**, **9–11**.

References

1. Jackson, S. E. and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2: 1. Evidence for a two-state transition. *Biochemistry* **30,** 10,428–10,435.
2. Li, A. and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **91,** 10,430–10,434.
3. Li, A. and Daggett, V. (1996) Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 using molecular dynamics simulations. *J. Mol. Biol.* **257,** 412–429.
4. Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1996) Structure of the transition state for folding of a protein derived from experiment and simulations. *J. Mol. Biol.* **257,** 430–440.

5. Daggett, V., Li, A., and Fersht, A. R. (1998) A combined molecular dynamics and ϕ-value analysis of structure-reactivity relationships in the transition state and pathway of barnase: the structural basis of Hammond and anti-Hammond effects. *J. Am. Chem. Soc.* **120,** 12,540–12,554.
6. Kazmirski, S. and Daggett, V. (1998) The structural and dynamic properties of unfolded proteins: simulations of the "molten coil" state of BPTI. *J. Mol. Biol.* **277,** 487–506.
7. Lazaridis, T. and Karplus, M. (1997) "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278,** 1928–1931.
8. Fulton, K. F., Main, E. R. G., Daggett, V., and Jackson, S. E. (1999) Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.* **291,** 445–461.
9. Alonso, D. O. V., Alm, E., and Daggett, V. (2000) Characterization of the unfolding pathway of the cell cycle protein P13suc1 via molecular dynamics simulations: implications for domain swapping. *Struct. Fold. Des.* **8,** 101–110.
10. Alonso, D. O. V. and Daggett, V. (2000) Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains. *Proc. Natl. Acad. Sci. USA* **97,** 133–138.
11. Daggett, V. and Levitt, M. (1992) A model of the molten globule state from molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **89,** 5142–5146.
12. Mark, A. E. and van Gunsteren, W. F. (1992) Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state. *Biochemistry* **31,** 7745–7748.
13. Tirado-Rives, J. and Jorgensen, W. L. (1993) Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* **32,** 4175–4184.
14. Alonso, D. and Daggett, V. (1995) Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J. Mol. Biol.* **247,** 501–520.
15. Storch, E. and Daggett, V. (1996) Structural consequences of heme removal: molecular dynamics simulations of rat and bovine apocytochrome b5. *Biochemistry* **35,** 11,596–11,604.

16. Laidig, K. E. and Daggett, V. (1996) Molecular dynamics simulations of apocytochrome b562: a "highly ordered" molten globule. *Folding and Design* **1,** 353–364.
17. Li, A. and Daggett, V. (1998) The unfolding of barnase: characterization of the major intermediate. *J. Mol. Biol.* **275,** 677–694.
18. Kazmirski, S. and Daggett, V. (1998) Nonnative interactions in protein folding intermediates: molecular dynamics simulation of hen lysozyme. *J. Mol. Biol.* **284,** 793–806.
19. Daggett, V. and Levitt, M. (1993) Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **232,** 600–619.
20. Daggett, V. (1993) A model for the molten globule state of CTF generated using molecular dynamics, in *Techniques in Protein Chemistry IV,* (Angeletti, R. H., ed.), Academic Press, San Diego, CA, pp. 525–532.
21. Vishveshwara, S., Ravishanker, C., and Beveridge, D. L. (1993) Differential stability of β-sheets and α-helices in β-lactamase: a high temperature molecular dynamics study of unfolding intermediates. *Biophys. J.* **65,** 2304–2312.
22. Caflisch, A. and Karplus, M. (1994) Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl. Acad. Sci. USA* **91,** 1746–1750.
23. Caflisch, A. and Karplus, M. (1995) Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **252,** 672–708.
24. Caflisch, A. and Karplus, M. (1999) Structure details of urea binding to barnase: a molecular dynamics analysis. *Struct. Fold. Des.* **7,** 477–488.
25. Hunenberger, P. H., Mark, A. E., and van Gunsteren, W. F. (1995) Computational approaches to study protein unfolding: hen egg white lysozyme as a case study. *Proteins Struct. Funct. Genet.* **21,** 196–213.
26. Tirado-Rives, J., Orozco, M., and Jorgensen, W. L. (1997) Molecular dynamics simulations of the unfolding of barnase in water and 8 *M* aqueous urea. *Biochemistry* **36,** 7313–7329.
27. Boczko, E. M. and Brooks, C. L. III (1995) First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269,** 393–396.

28. Guo, Z., Brooks, C. L., III, and Boczko, E. M. (1997) Exploring the folding free energy surface of a three-helix bundle protein. *Proc. Natl. Acad. Sci. USA* **94,** 10,161–10,166.
29. Sheinerman, F. B. and Brooks, C. L. (1998a) Calculation on folding of segment B1 of Streptococcal protein G. *J. Mol. Biol.* **278,** 439–456.
30. Sheinerman, F. G. and Brooks, C. L. (1998b) Molecular picture of folding of a small α/β protein. *Proc. Natl. Acad. Sci. USA* **95,** 1562–1567.
31. Bond, C. J., Wong, K., Clarke, J., Fersht A. R., and Daggett, V. (1997) Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. *Proc. Natl. Acad. Sci. USA* **94,** 13,409–13,413.
32. Wong, K., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M. V., Fersht, A. R., et al. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and their role in folding. *J. Mol. Biol.* **296,** 1257–1282.
33. Alonso, D. O. V. and Daggett, V. (1998) Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci.* **7,** 860–874.
34. Duan, Y. and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282,** 740–744.
35. Daura, X., Jaun, B., Seebach, D., van Gunsteren, W. F., Mark, A. E. (1998) Reversible peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.* **280,** 925–932.
36. Alonso, D. O. V., Pan, Y. P., Ham, S., Bennion, B., and Daggett, V. (2000) Increasing temperature accelerates unfolding without changing the pathway of unfolding or the protein energy landscape. (In preparation).
37. Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. L. (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340,** 122–126.
38. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224,** 771–782.
39. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995) Movement of the position of the transition state in protein folding. *Biochemistry* **34,** 13,656–13,662.
40. Matouschek, A. and Fersht, A. R. (1993) Application of physical organic chemistry to engineered mutants of proteins: hammond pos-

tulate behavior in the transition state of protein folding. *Proc. Natl. Acad. Sci. USA* **90,** 7814–7818.

41. Serrano, L., Matouschek, A., and Fersht, A. R. (1992) The folding of an enzyme III. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* **224,** 805–818.
42. Matthews, J. M. and Fersht, A. R. (1995) Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for the structural elements for unfolding/folding of barnase. *Biochemistry* **34,** 6805–6814.
43. Jackson, S. E., el Masry, N., and Fersht, A. R. (1993) Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. *Biochemistry* **32,** 11,270–11,278.
44. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor-2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254,** 260–288.
44. Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., Prat Gay, G., and Fersht, A. R. (1995) Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. *J. Mol. Biol.* **254,** 289.
45. Otzen, D. E., Itzhaki, L. S., el Masry, N. F., Jackson, S. E., and Fersht, A. R. (1994) Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implication for mechanisms of protein folding. *Proc. Natl. Acad. Sci. USA* **91,** 10,422–10,425.
46. Fersht, A. R., Itzhaki, L. S., el Masry, N. F., Matthews, J. M., and Otzen, D. E. (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transtion state. *Proc. Natl. Acad. Sci. USA* **91,** 10,426–10,429.
47. Nölting, B., Golbik, R., Neira, J. L., Soler-Gonzalez, A. S., Schreiber, G., and Fersht, A. R. (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. USA* **94,** 826–830.
48. Lopez-Hernandez, E. and Serrano, L. (1996) Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Folding and Design* **1,** 43–55.

49. Villegas, V., Martinez, J. C., Aviles, F. X., and Serrano, L. (1998) Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain *J. Mol. Biol.* **283,** 1027–1036.
50. Viguera, A. R., Serrano, L., and Wilmanns, M. (1996) Different folding transition states may result in the same native structure. *Nat. Struct. Biol.* **3,** 874.
51. Martinez, J. C., Pisabarro, M. T., and Serrano, L. (1998) Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat. Struct. Biol.* **5,** 721.
52. Grantcharova, V. P., Riddle, D. S., Santiago, J. V., and Baker, D. (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.* **5,** 714–720.
53. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry* **37,** 5337–5343.
54. Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M., and Baker, D. (1998) The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* **284,** 807–815.
55. Schymkowitz, J. W. H., Rousseau, F., Irvine, I. R., and Itzhaki, L. S. (2000) Structure of the transition state and major intermediate for the folding of the cell cycle regulatory protein p13suc1 analysed using protein engineering. *Struct. Fold. Des.* **8,** 89–100.
56. Dobson, C. M. (1992) Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* **2,** 6–12.
57. Shortle, D. R. (1996) Structural analysis of nonnative states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* **6,** 24–30.
58. Wüthrich, K. (1994) NMR assignments as a basis for structural characterization of denatured states of globular proteins. *Curr. Opin. Struct. Biol.* **4,** 93–99.
59. Neri, D., Billeter, M., Wider, G., and Wüthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434 repressor. *Science* **257,** 1559–1563.
60. Alexandrescu, A. T. and Shortle, D. (1994) Backbone dynamics of a highly disordered 131 residue fragment of Staphylococcal nuclease. *J. Mol. Biol.* **242,** 527–546.
61. Alexandrescu, A. T., Abeygunawordana, C., and Shortle, D. (1994) Structure and dynamics of a denatured 131-residue fragment of Sta-

phylococcal nuclease: a heteronuclear NMR study. *Biochemistry* **33,** 1063–1072.

62. Arcus, V. L., Vuilleumier, S., Freund, S. M., Bycroft, M., and Fersht, A. R. (1994) Toward solving the folding pathway of barnase: the complete backbone ^{13}C, ^{15}N and ^{1}H NMR assignments of its pH-denatured state. *Proc. Natl. Acad. Sci. USA* **91,** 9412–9416.
63. Arcus, V. L., Vuilleumier, S., Freund, S. M., Bycroft, M., and Fersht, A. R. (1995) A comparison of the pH, urea and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J. Mol. Biol.* **254,** 305–321.
64. Logan, T. M., Thériault, Y., and Fesik, S. W. (1994) Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.* **236,** 637–648.
65. Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci.* **4,** 2605–2615.
66. Pan, H., Barbar, E., Barany, G., and Woodward, C. (1995) Extensive nonrandom structure in reduced and unfolded bovine pancreatic trypsin inhibitor. *Biochemistry* **34,** 13,974–13,981.
67. Wong, K. B., Freund, S. M. V., and Fersht, A. R. (1996) Cold denaturation of barstar: ^{1}H, ^{15}N and ^{13}C NMR assignment and characterization of residual structure. *J. Mol. Biol.* **259,** 805–818.
68. Zhang, O. and Forman-Kay, J. D. (1995) Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry* **34,** 6784–6794.
69. Zhang, O. and Forman-Kay, J. D. (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry* **36,** 3959–3970.
70. Wang, Y. and Shortle, D. (1995) The equilibrium folding pathway of staphylococcal nuclease: identification of the most stable main-chain interactions by NMR and CD spectroscopy. *Biochemistry* **34,** 15,895–15,905.
71. Wang, Y. and Shortle, D. (1996) A dynamic bundle of four adjacent hydrophobic segments in the denatured state of Staphylococcal nuclease. *Protein Sci.* **5,** 1898–1906.
72. Freund, S. M. V., Wong, K., and Fersht, A. R. (1996) Initiation sites of protein folding by NMR analysis. *Proc. Natl. Acad. Sci. USA* **93,** 10,600–10,603.

73. Gillespie, J. R. and Shortle, D. (1997a) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268,** 158–169.
74. Gillespie, J. R. and Shortle, D. (1997b) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268,** 170–184.
75. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., et al. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 *M* urea. *Biochemistry* **36,** 8977–8991.
76. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1995) Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry* **34,** 868–878.
77. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, et al. (1997) NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *J. Mol. Biol.* **274,** 152–159.
78. Fong, S., Bycroft, M., Clarke, J., and Freund, S. M. V. (1998) Characterisation of urea-denatured states of an immunoglobulin superfamily domain by heteronuclear NMR. *J. Mol. Biol.* **278,** 417–429.
79. Meekhof, A. E. and Freund, S. M. V. (1999) Probing residual structure and backbone dynamics on the milli- to picosecond timescale in a urea-denatured fibronectin type III domain. *J. Mol. Biol.* **286,** 579–592.
80. Panchenko, A. R., Luthey-Schulten, A., and Wolynes, P. G. (1996) Foldons, protein structural modules, and exons. *Proc. Natl. Acad. Sci. USA* **93,** 2008.
81. Li, A. and Daggett, V. (1995) Investigation of the solution structure of chymotrypsin inhibitor 2 using molecular dynamics: comparison to X-ray and NMR data. *Protein Eng.* **8,** 1117–1128.
82. Shaw, G. L., Davis, B., Keeler, J., and Fersht, A. R. (1995) Backbone dynamics of chymotrypsin inhibitor 2: Effect of breaking the active site bond and its implications for the mechanism of inhibition of serine proteases. *Biochemistry* **34,** 2225–2233.
83. Itzhaki, L. S., Neira, J. L., and Fersht, A. R. (1997) Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J. Mol. Biol.* **270,** 89–98.

84. Killick, T. R., Freund, S. M. V., and Fersht, A. R. (1998) Real-time NMR studies on folding of mutants of barnase and chymotrypsin inhibitor 2. *FEBS Lett.* **423,** 110–112.
85. Oliveberg, M., Tan, Y-J., Silow, M., and Fersht, A. R. (1998) The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* **277,** 933–943.
86. Ladurner, A. G., Itzhaki, L. S., Daggett, V., and Fersht, A. R. (1998) Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. USA* **95,** 473–8478.
88. Prat Gay, G., Ruiz-Sanz, J., Davis, B., and Fersht, A. R. (1994) The structure of the transition state for the association of two fragments of the barley chymotrypsin inhibitor 2 to generate native-like protein: implications for mechanisms of protein folding. *Proc. Natl. Acad. Sci. USA* **91,** 10,943–10,946.
89. Neira, J. L., Itzhaki, L. S., Ladurner, A. G., Davis, B., Prat Gay, G., and Fersht, A. R. (1997) Following co-operative formation of secondary and tertiary structure in a single protein module. *J. Mol. Biol.* **268,** 185–197.
90. Tan, Y.-J. (1995) Phil.D. Thesis, Cambridge University.
91. Kazmirski, S., Wong, K., Freund, S. M. V., Tan, Y. J., Fersht, A. R., and Daggett, V. (2001) Protein folding from a highly disordered denatured state. (In preparation).
92. Tan, Y.-J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pKa-values in the denatured states of proteins. *J. Mol. Biol.* **254,** 980–992.
93. Karplus, M. (1959) Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **30,** 11–15.
94. Shakhnovich, E. U., Abkevich, V., and Ptisyn, O. (1996) Conserved residues and the mechanism of protein folding. *Nature* **379,** 96–98.
95. Ladurner, A. G., Itzhaki, L. S., and Fersht, A. R. (1997) Strain in the folding nucleus of chymotrypsin inhibitor 2. *Folding and Design* **2,** 363–368.
96. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) The MIDAS display system. *J. Mol. Graphics.* **6,** 13–27.
97. Kraulis, P. J. (1991) MolScript: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* **24,** 946–950.

Index

METHODS IN MOLECULAR BIOLOGY™ • 168
Series Editor: *John M. Walker*

Protein Structure, Stability, and Folding

Edited by

Kenneth P. Murphy

University of Iowa College of Medicine, Iowa City, IA

With the huge increase in available data on the DNA sequences of proteins, there is a growing need to understand and characterize how proteins fold into their biologically active native states and the basis for the stability of these states. In *Protein Structure, Stability, and Folding,* Kenneth P. Murphy and a panel of internationally recognized investigators describe some of the newest experimental and theoretical methods for investigating these critical events and processes. Among the techniques discussed are the many methods for calculating aspects of protein stability and dynamics from knowledge of the structure, for calculating conformational entropy, and for performing molecular dynamics simulations of protein unfolding. New experimental approaches presented include the use of co-solvents, novel applications of hydrogen exchange techniques, temperature-jump methods for looking at folding events, and new strategies for mutagenesis experiments.

Unique in its powerful combination of theory and practice, *Protein Structure, Stability, and Folding* offers protein and biophysical chemists the means to gain a more comprehensive understanding of this complex area by detailing many of the major innovative techniques in use today.

FEATURES

- **Provides an extensive collection of innovative experimental and theoretical techniques**
- **Emphasizes the theory behind each experimental approach**
- **Details new applications for hydrogen exchange techniques**
- **Includes cutting-edge methods for calculating conformational entropy**

CONTENTS

ISBN 0-89603-682-0
90000>
9 780896 036826

Methods in Molecular Biology™ • 168
PROTEIN STRUCTURE, STABILITY, AND FOLDING
ISBN: 0-89603-682-0
http://humanapress.com