Methods in Molecular Biology™

VOLUME 167

DNA Sequencing Protocols

SECOND EDITION

Edited by

Colin A. Graham

Alison J. M. Hill

HUMANA PRESS

2

The Universal Primers and the Shotgun DNA Sequencing Method

Joachim Messing

1. Introduction

For studies in molecular biology, DNA purification has been essential, in particular for DNA sequencing, probing, and mutagenesis. The amplification of DNA in *Escherichia coli* by cloning vehicles derived from Ml3mp or pUC made expensive physical separation techniques such as ultracentrifugation unnecessary. Although today the polymerase chain reaction (PCR) is a valuable alternative for the amplification of small DNA pieces ***(1)***, it cannot substitute for the construction of libraries of DNA fragments. Therefore, *E. coli* has served not only as a vehicle to amplify DNA, but also to separate many DNA molecules of similar length and the two DNA strands simultaneously. For this purpose, a bacteriophage such as M13 can be used. The various viral *cis*- and *trans*-acting functions are critical not only for strand separation, but also to separate the single-stranded DNA from the *E. coli* cell by an active transport mechanism through the intact cell wall.

Although it may have been somewhat surprising to some how many changes in its DNA sequence the phage tolerated, manipulations of this amplification and transport system have been extended

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

today, even to the viral coat proteins for the production of epitope libraries ***(2)***. Much of the work is now more than a decade old, but experience has confirmed the usefulness of some simple biological paradigms. Techniques that were new and limiting 20 years ago included automated oligonucleotide synthesis and the use of thermostable enzymes, which add a critical dimension to molecular biology today. Neither necessarily replaces the previous techniques, but they create greater flexibility, enormously accelerate our investigations, and even make certain analyses possible for the first time.

2. Emergence of DNA Synthesis Applications

In 1974, at one of the first meetings on the use of restriction endonucleases in molecular biology DNA synthesis and sequencing methods had broken new grounds. Work on the chemical synthesis of a *tRNA* gene was presented, and the initial work on sequencing phage ΦX174 using restriction fragments as primers for the so-called plus–minus method was discussed. At that time, oligonucleotide synthesis required a major effort and could not be applied easily in general. Restriction fragments offered an alternative. They could be used as primers for DNA synthesis in vitro and for marker-rescue experiments to link genetic and physical maps of viruses such as SV40, both forerunners for DNA sequencing and site-directed mutagenesis.

There were several reasons to use ΦX174 as the first model in developing DNA-sequencing techniques and determining the sequence of an entire autonomous genome. First, it was one of the smallest DNA viruses; it is even smaller than M13. Second, the mature virus consists of single-stranded DNA, eliminating the need to separate the two strands of DNA for template preparation; this is even more critical if one wishes to use double-stranded restriction fragments as primers. Third, a restriction map was superimposed on the genetic map by marker-rescue experiments ***(3)***. Physical mapping still serves today as a precondition for sequencing other genomes. Restriction sites were critical as signposts along the thousands of nucleotides and provided the means to dissect the double-

stranded replicative form or RF of ΦX174 in small but ordered pieces that permitted the DNA sequencing effort to proceed in a walking manner along the genome. Today the restriction map can be replaced by any DNA sequence (e.g., STS) because the synthesis of oligonucleotides is so rapid that we can use the DNA sequence that was just read from a sequencing gel to design and produce an oligonucleotide to extend the sequencing gel further in the 5' direction. Therefore, the use of oligonucleotides instead of restriction fragments in such a primer walking method would have enormously accelerated the ΦX174 project.

At the Cleveland Conference on Macromolecules in 1981, after a talk that I had given, the replacement of shotgun sequencing by such a method was suggested by Marvin Caruthers, who, as a pioneer in DNA synthesis and its automation ***(4)***, saw a perfect match of this emerging technology with DNA sequencing. Another expert in the chemical synthesis of DNA, Michael Smith, recognized the potential of the chemistry as a DNA mutagen ***(5)***. Design of oligonucleotides as mutagens, however, requires the knowledge of the target DNA sequence. Therefore, the location of amber mutations in ΦX174 genome and the use of restriction fragments to rescue them was critical not only to delineate a physical and genetic map for DNA sequencing, but also to develop oligonucleotide site-directed mutagenesis ***(3)***. It is clear that today's protein engineering had its roots right there with the right people at the right time. It was also possible to show that oligonucleotides eliminated the need for strand separation for DNA sequencing ***(6)***. Even double-stranded DNA sequencing with universal primers became easier with the development of the pUC plasmids ***(7)***. However, application of oligonucleotide site-directed mutagenesis to cloned DNA was also greatly accelerated by the development of single-stranded DNA cloning vectors ***(8,9)***.

Despite all the advantages of choosing ΦX174 as a model system, Sanger's group nearly picked a different single-stranded DNA phage, fd. In principle, *E. coli* has two different types of single-stranded DNA phage, represented by ΦX174 and fd. The first is packaged into an icosahedral head; it kills and lyses the host cell,

but does not require F pili, which are receptor sites on the surface of the cell wall encoded by F factors. Its host range is restricted to *E. coli* C. Phage such as fd can infect only male-specific *E. coli* producing pili at their surface that are packaged in a filamentous coat and discharged from the cell without lysis; infected cells can continue to divide. These differences are critical, but there was another reason for choosing fd originally. The major coat protein encoded by gene *VIII* of the phage, a very small but very abundant protein, had been sequenced by protein sequencing methods. Therefore it seemed obvious, particularly to someone who had pioneered protein sequencing, to use protein sequence to check the DNA sequence. The protein sequence allowed the design and synthesis of an oligonucleotide that would prime in vitro DNA synthesis within the coat protein gene. Furthermore, the derived DNA sequence had to match the protein sequence. Of course, the codon redundancy of many amino acids made it difficult to design a unique primer, and it might not have been too surprising that the approach did not lead to the correct DNA sequence ***(10)***. It turned out later that this was owing not to the choice of codons, but to a mistake in the protein sequence. Still, cloning based on protein sequence information again has its roots right there.

3. Replication Systems and M13

In 1974, there was a great interest in understanding how DNA is replicated. Viruses and plasmids were used to identify components of the cell machinery required for the initiation and elongation process of DNA replication. These studies were also pursued at the "Abteilung" of Virology at the Max Planck Institute of Biochemistry in Munich. Eleven years earlier, Hofschneider isolated a filamentous phage from the Munich sewers that he named after a series of phage with the initial M ***(11)***. Number 13 was the one that was studied most. Looking for a different research topic than DNA replication, I explored the possibility of combining M13 phage production with the in vitro DNA synthesis-based method of DNA sequencing.

Although this might have been obvious to those familiar with phage replication, innovative methods were needed for adaptation to DNA cloning techniques. The walking method for sequencing ΦX174 was the strategy used at that time, and I thought that it would be difficult to clone large fragments into M13 (although the author's record was approx 40 kb) and that a walking method might therefore have a limited use. Logically, the only alternative to the walking method was the use of shotgun cloning and a universal primer. The replicative form of M13 could be used to clone DNA fragments of a size slightly larger than necessary for single sequencing reactions, and a universal sequence near the cloning site would be used as a primer. This would shift the work from preparing primers to preparing templates, which has become essential for high-throughput DNA sequencing ***(12)***. If they were numerous, cloning was much faster than any biochemical technique. In view of these considerations, a plan took shape to construct in vitro recombinants of phage M13 using a different method from any other cloning methods. One might recall that in vitro recombinants were usually based on drug-resistance markers. This led to the development of plasmid vectors with unique cloning sites that were scattered all over the plasmid genome ***(13)***. However, such an approach is not compatible with the life cycle of M13.

4. New Mutants of Filamentous Phage

Both Schaller's and Zinder's laboratory considered, and later used, transposons to develop f1 and fd transducing phage ***(14,15)***. However, one could predict that such a course of experiments, although useful for plasmid cloning vehicles, would be less useful for M13. It appeared plausible, and such an experiment could demonstrate that, in contrast to ΦX174, filamentous phage can accommodate additional DNA by extending the filamentous coat; infected cells can be treated like plasmid-containing cells. To some degree this had already been proven since phage mutants of more than unit length had been described ***(16)***. Another advantage of transposon mutagenesis was that insertion mutants would be naturally selected.

Selection of an insertion site in the phage genome was one of the greatest obstacles from the beginning. Although plasmids and bacteriophage λ were natural transducing elements, filamentous phage had never been shown to have this property, and it was not obvious whether insertion mutants would be viable. This was difficult particularly because it was already known that amber mutants of most viral genes not only cause abortive infection, but also lead to killing of the host, which does not occur when these mutations are suppressed. Therefore, insertion mutants that are treated like plasmids would kill the cell.

Thus, it was predictable that insertion mutants had to be restricted to noncoding regions. Therefore, a decision was made to use a restriction enzyme that recognized at least two different sites in the intergenic region of the RF. Rather than asking whether the intergenic region contains a target site for transposons, it was decided to look at the restriction map, which showed that it was possible to obtain at least two different insertion mutants within the intergenic region. The only difficulty in such an experiment was to find conditions where the restriction enzyme would cut RF only once but at any of the possible target sites, so that a population of unit length RF could be ligated to the appropriate marker DNA fragment.

However, there was another reason not to use drug-resistance markers. Infected cells still divide, but very slowly. Therefore, selection takes much longer than with plasmids, but it makes it very easy to distinguish infected from noninfected cells on a bacterial lawn. A single infection grown on a bacterial lawn forms a turbid plaque. If bacterial cells are transfected by the calcium chloride technique of Mandel and Higa ***(17)***, a transformed cell can be recognized as a plaque. Hence, no selection technique is necessary. Still, how would one be able to distinguish between wild-type M13 and M13 insertion mutants? Although it was quite plausible to think of the histochemical screen used for bacteriophage λp*lac* by Malamy et al. ***(18)***, the *lacZ* gene would have been a large insertion. However, rather than using entire genes as markers, one could clone only the portion encoding the amino (N)-terminal and the repressible control region and provide the rest *in trans* by the host of the phage.

This became clear when Landy et al. ***(19)*** wrote on the purification of a 800 bp *Hin*dII fragment from λp*lac* capable of α-complementation in a cell-free transcription–translation system.

An informal sequence of this fragment showed that it was 789 bp long and included the first 146 codons of the *lacZ* gene, but it was necessary to assemble many components and purify several restriction endonucleases. Work began after some strains and purified *lac* repressor were traded: This allowed the purification of the 789 bp *Hin*dII fragment out of about 50 other restriction fragments by simply filtering the DNA–*lac* repressor complex through a nitrocellulose filter. After adding isopropylthiogalactoside (IPTG), it was possible to recover the DNA in solution. Using DNA binding proteins for purifying and cloning promoter regions has become a very important technique today. Another well-established procedure, the ligation of restriction fragments via blunt ends, was also untested when this experiment was ready. This explains why only two transformants were obtained ***(20)***—one of them was saved and named Ml3mpl (mp for Max Planck Institute, M for Munich). Electron microscopy proved that added DNA was packaged as a filamentous phage and produced as single-stranded DNA ***(21,22)***.

Now the path took a more formal shape. Not only would the histochemical screen work by detecting a blue among colorless plaques, but it could also be reversed. One uncertainty was how to introduce new restriction sites in the right region. Such a site had to be unique for Ml3mpl and positioned not somewhere in the viral genome, but in the *lacZ* region, so that insertion mutants would not give rise to blue plaques. Inspection of the sequence showed that there were not many sequences in the N-terminal region that could be converted in a single step into a unique restriction site. Attempts to use *Eco*RI linkers that became available at the time to "marker rescue" them did not succeed, probably because they were too short. Without somebody to synthesize a customized oligonucleotide homologous to the *lac* region, it was fruitless to continue this approach. As an alternative, a chemical mutagen seemed to be more practical. It was known that methylated G could mispair with uracil or thymine. Therefore, by methylating the single-stranded M13

The first ten codons of the *lacZ* gene

```
1   2   3   4   5   6   7   8   9   10  M13mp1
ATG ACC ATG ATT ACG GAT TCA CTG GCC GTC (+ or viral strand)
                  GGmAT TC   (+ or viral strand)
                  CT TA AG   (- or complementary strand)

1   2   3   4   5   6   7   8   9   10  M13mp2
ATG ACC ATG ATT ACG AAT TCA CTG GCC GTC (+ or viral strand)
                  EcoRI
```

Fig. 1. Creation of an *Eco*RI site by chemical mutagenesis. By screening the nucleotides of the first 10 codons of the *lacZ* gene, we found that the sequence GGATTC could be converted into either an *Eco*RI site GAATTC or a *Bam*HI site GGATCC by a single base change. Because a *Bam*HI site was already present in gene III, but no *Eco*RI site in M13mp1, and I had an ample supply of *Eco*RI enzyme purified in my laboratory, we decided to select the GAATTC site, which also changed codon GAT for aspartic acid to AAT for asparagine ***(23)***.

DNA with nitrosomethylurea, a mutation could be introduced into the minus strand and the subsequent RF molecules (**Fig. 1**).

Unfortunately, there was no good genetic selection for this procedure. It would require brute force methods of enriching *Eco*RI-sensitive RF from a transformed phage library by gel electrophoresis of linear vs circular molecules. Still, it was difficult to believe at the time when a mutant M13 RF was isolated that was not only sensitive to *Eco*RI, but also had exactly the predicted base change in codon 5 of the *lacZ* gene ***(23)***. The same mutagenesis led to two more *Eco*RI mutants and a mutant RF that was resistant to *Bam*HI, a site within gene *III*. At that time, there was still much concern that many mutations might not be tolerated because of changes in the protein sequence or the secondary structure of RNA, but this *Bam*HI site served later on as a cloning site for epitope libraries ***(2)***. Still, changes in the *lac* DNA should probably occur at a higher frequency than in the viral DNA. Furthermore, the N-terminus of the *lacZ* gene appeared to be more flexible because it was demonstrated that fusion proteins retained β-galactosidase function. On the other hand, the

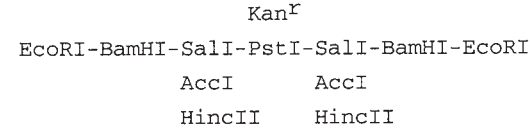
tremendous selection power for suppressor mutations of phage insertion mutants remained elusive for many years. In other words, any mutation that was introduced in one location of the viral genome could potentially be compensated for by a second mutation somewhere else. The primary mutant might give a low titer, but because of the growth advantage a suppressor mutant would take over rapidly. An example of such a case is Ml3mpl. Dotto and Zinder ***(24)*** showed that insertion mutants at the mpl *Hae*III site gave a low titer phenotype. Since Ml3mpl gave a normal titer, they searched for a suppressor mutation. Codon 40 in gene *II* of Ml3mpl indeed was changed.

5. Polycloning Sites, a Universal Primer, and Biosafety

Using chemical mutagens other restriction sites were eliminated. The reason for this becomes clear by considering the elements of the enzymatic DNA sequencing technique ***(25)***. The *Eco*RI site in Ml3mp2 allowed cloning by screening for colorless plaques and prepare a template for sequencing the inserted DNA. Still, adjacent *lac* DNA needed to be purified as a primer, and initially a small restriction fragment was subcloned into a plasmid for primer purification ***(26)***. However, such a primer fragment needed to be denatured because it was double-stranded, and to be cut off after the sequencing reaction to produce a shift of the 3' end in the sequencing gel.

Although such a protocol could still be improved on, a more serious obstacle arose suddenly from the concern over the biological containment of M13 recombinant DNA. The NIH Recombinant Advisory Committee or RAC thought that the conjugation proficient *E. coli* host strains could lead to the spread of M13 infection and pose a risk in using M13 as a cloning vehicle. On the other hand, using one of the *traD* or *traI* mutants reduces conjugation by a factor of 10^6, leaving the infection of M13 unabated. Because the F factor carrying the *traD* mutation was wild-type *lac,* a histochemical screen with the mp vectors would not be possible and one would have to return to drug-resistant-type M13 vectors. The scientific reasoning of RAC is hard to understand. First, nobody argued against

A Polylinker of pUC4K:



1) Partial digest of target DNA with *Mbo*I, *Hpa*II, and *Taq*I for instance.
2) Cut pUC4K with *Bam*HI or *Acc*I to completion.
3) Clone Kan^r into target DNA.
4) Cut target DNA with *Pst*I and select for Kan^s.
5) Insertions are either insertions of eight (*Bam*HI) or four (*Acc*I) amino acids.

Fig. 2. Symmetric **(A)** and asymmetric **(B)** *(opposite page)* polylinkers.

Two other "tricks" were used in the construction of polylinkers. One type of polylinker was symmetrical, in which all sites except the central one occurred twice. By cloning a drug-resistance marker into the central site, the polylinker could be used in a linker scanning method of coding regions ***(7)*** and ***(35)***. The other type of polylinker was a pair, where two vectors contain an array of sites only once, but each of them in the opposite orientation. Cloned DNA no longer could be cloned out with a single enzyme as in the first type, but DNA could be cloned by using two different sites at the same time. This had the advantage that the orientation of a cloned fragment could be determined. By using a vector pair, both orientations can be obtained with the same pair of restriction cuts and therefore each strand of a restriction fragment could become the viral strand of M13 and available as a template for sequencing. Furthermore, by using two restriction enzymes that produce 3' and 5' overhangs, one can use it either for cloning oligonucleotide libraries or to generate unidirectional deletions with exonuclease III.

Agrobacterium tumefaciens as a plant transformation vector, although it was conjugation proficient and easily could spread in the environment. Second, F pili were never made under stress or anaerobic conditions, something that was already known as "phenocopies." Conjugation in the human gut was in any case nearly zero. Third, M13 infection *per se* reduces conjugation by a factor of 10^6. Leaving these arguments unanswered, a new series of *E. coli* strains (JM series) all carrying the M15 deletion on the F' *traD*36

B Polylinker of M13mp8 and M13mp9

```
EcoRI-SmaI-BamHI-SalI-PstI-HindIII  (mp8)
      XmaI        AccI
                  HincII

HindIII-PstI-SalI-BamHI-SmaI-EcoRI  (mp9)
             AccI        XmaI
             HincII

5'----->3' GATCXXXXXXXXXXXXXXXXXXXTGCA
GAATTCCCGGG     <---- exoIII           GCCAAGCTT
CTTAAGGGCCCCTAG  <-- exoVII----> ACGTCGGTTCGAA
          BamHI              PstI
```

1) The small internal fragment is lost and recircularization (eliminating the background of recircularized vector) is only possible by providing the target DNA with the correct sticky ends.
2) It can only be cloned in one orientation, for the other orientation the other vector has to be used.
3) The notch can be used to clone oligonucleotide libraries with the first and the last four nucleotides in common.
4) Since exonuclease III requires double-stranded DNA and exonuclease VII single-stranded DNA, one can be used to attack the 3' recessed end of the *Bam*HI site, and the other both, the single-stranded ends of *Bam*HI and the *Pst*I sites.

Fig. 2B.

episome was constructed. Because it was a concern of NIH, a publication circulated to all potential users by NIH was used to describe the construction of the JM series and list the primer plasmids and the different M13 vectors, even already one with a *Hin*dIII site flanked by *Eco*RI sites that could produce still a blue plaque ***(27)***.

Although sequencing of eukaryotic DNA by M13 cloning was now possible, preparation of the primer from the plasmid was still cumbersome. It was clear that an oligonucleotide was needed to replace the restriction fragment as a universal primer. Although commercially custom-synthesized oligonucleotides were still very

expensive, the emergence of new chemical DNA synthesis methods such as the "triester method" allowed the demonstration of a short oligonucleotide as a universal primer ***(28)***. However, oligonucleotide synthesis had other applications as well.

In 1978, another interesting observation was made, namely that inframe insertions of linkers in the *Eco*RI site could still give a positive color reaction. One of these isolates, Ml3mp5, already listed in the NIH Bulletin described earlier ***(27)***, could be used to clone both *Eco*RI and *Hin*dIII fragments at the same site and with the same primer for sequencing. The utility of creating cloning sites on top of each other was based on the universal primer concept, but in turn caused the development of multiple cloning sites (MCS) or polylinkers that are found in all cloning vehicles today and provide many additional uses (**Fig. 2**). Therefore, work began on the synthesis of an oligonucleotide that could be inserted into the *Eco*RI site and generate restriction sites recognized by six base pair cutters such as *Bam*HI, *Acc*I, *Sma*I, or *Hin*dIII useful for cloning either blunt-ended fragments or fragments with sticky ends produced by four base cutters such as *Sau*3A, *Taq*I, and *Hpa*II (**Fig. 3**) (consistent with a DNA shotgun sequencing approach). This also required a renewed chemical mutagenesis to eliminate the *Acc*I and the *Hinc*II sites naturally occurring in M13. All single mutations were combined by marker rescue to give rise to Ml3mp7. Another important change in the enzymatic DNA sequencing protocol was required, facilitating the handling of many templates at the same time. Although the original sequencing method was carried out in sealed glass capillaries ***(25)***, reactions were now carried out in Eppendorf tubes ***(28)***.

6. Shotgun Sequencing in Practice

This was just the system, but did shotgun sequencing succeed? Initially, lack of funds and proper laboratory facilities made a demonstration impossible. Financial support finally arrived from the USDA, and the entire genome (8031 bp) of a plant virus, cauliflower mosaic virus (CaMV), was determined by DNA shotgun sequenc-

Unique vector sequence	Compatible target sites
G'GATC C	N'GATC N
C CTAG'G	N CTAG'N
*Bam*HI	*Sau*3A, *Mbo*I
GT'CG AC	NT'CG AN
CA GC'TG	NA GC'TN
*Acc*I	*Taq*I
	NC'CG GN
	NG GC'CN
	*Hpa*II
GTC'GAC	NNN'NNN
CAG'CTG	NNN'NNN
*Hinc*II	Restriction enzymes like *Alu*I, *Hae*III, etc.; *Bal*31 or *Exo*III/V; sheared and repaired DNA

Fig. 3. Sticky and blunt-end cloning of small fragments into unique cloning sites. By designing a unique sequence for the M13mp vectors that were recognized by restriction enzymes that could cut a hexanucleotide sequence in various ways by either producing sticky ends of four or two bases, or blunt ends, the variety of DNA fragments that could be cloned next to the universal primers were endless. Note that the sequence GTCGAC was recognized by *Sal*I, *Acc*I, and *Hinc*II, each producing different ends. In our sequencing project with cauliflower mosaic virus, we generated small DNA fragments for shotgun DNA sequencing by cleaving CAMV with *Eco*RI*, *Mbo*I, *Hpa*II, *Taq*I, *Hinc*II, *Hae*III, and *Alu*I ***(29)***. Later, we used DNase I ***(36)***, sonication ***(31)***, and a combination of exonuclease III and VII to generate blunt ends ***(37)***.

ing for less than $3 a base. This was accomplished in a record time of 3 mo and finally published 1 yr later ***(29)***.

During the same time, work was progressing in sequencing human mitochondrial DNA twice the length of CaMV, but came to a grinding halt because the British guidelines emulated the NIH guidelines limiting the use of M13 to hosts with conjugation-reduced F factors. Because the Ml3mp2 vector and the universal

primer were already used in this project, both the JM strains and the newly developed Ml3mp7 for blunt-end cloning presented a major turning point for this work. Still, it seems hard to believe that a manuscript describing the concept of M13 shotgun DNA sequencing was rejected in the course of a review by PNAS as trivial at the time. However, it became clear that a few publications could not prove what many laboratories subsequently have proven. Therefore, it was critical not to have restricted the dissemination of the M13/pUC system (e.g., patent restrictions), but to have used the entire scientific community as a laboratory at large. This indeed became reality when this work became the most frequently cited work from 1981 to 1990 ***(30)***.

Becoming overwhelmed by requests for strains and protocols, the newly developing reagent companies were turned to for help. Their educational and service role immensely helped to disseminate the knowledge needed to train students and investigators in academia and industry in M13/pUC cloning, sequencing, and site-directed mutagenesis ***(31)***. Along the way, the first Apple-based software on shotgun sequencing ***(32,33)*** in addition to an undergraduate textbook ***(34)***. A good overview of Ml3mp, pUC vectors, and helper phage has also appeared ***(20)***.

Clearly, there are many new innovations that followed and made DNA sequencing what it is today. The substitution of X-ray films by lasers in conjunction with fluorescently labeled nucleotides allowed the automation of base calling ***(38)***. The throughput of templates per gel received improvement of fourfold, when all four chain terminators were synthesized with different dyes and could be separated in a single lane of a polyacrylamide gel ***(39)***. Subsequently, tracking software for these gels improved as well so that today reactions from 96 templates can be analyzed simultaneously (e.g., ABI 377, Applied Biosystems, Foster City, CA). Preparing gels and the running time of each gel would, however, limit the throughput of megabases of DNA sequences. The recent replacement of gels by capillaries yet provides another boost to the throughput of templates for sequencing. A major time-consuming step has been the loading of sequencing reactions on a gel. However, capillaries are loaded

with a robot arm from an array of reactions in 96-well microtiter plates. Because separation of DNA chains in capillaries is much faster than through gels, a station with four plates can keep the machine running in an uninterrupted fashion (e.g., ABI 3700, Applied Biosystems, Foster City, CA).

Calling bases from an autoradiogram has provided an important quality control that was lost when it was automated. Therefore, it was important to develop new computer software to control the accuracy of sequence information directly from the fluorograms produced by the DNA sequencing machines instead of their output of sequences ***(40,41)***. Without such a screen, the assembly of overlapping sequence reads becomes inefficient. Moreover, the faster throughput of sequencing reaction from shotgun libraries presented a major challenge to assemble the complete sequence from a set of overlapping fragments. The assembly of the 1.83 Mb *Haemophilus influenzae* Rd genome from more than 10^4 sequence reads has provided a good test case ***(12)***. Fragments of 1.6–2 kb were cloned into pUC18 instead of M13mp18. As shown previously, it permits the sequencing of an insert from both ends with universal primers ***(7)***. The advantage of having sequences from both ends aid in the computational assembly of fragments because their distances are known.

While all these improvements have accelerated the output of sequence information, one original feature has become even more prominent. The shotgun sequencing method sought to replace physical mapping by sequencing. Restriction maps, the hallmark of physical maps, could easily be reconstructed from the DNA sequence. By sequencing, restriction mapping became instant. For sequencing genomes larger than bacteria, however, bacterial artificial chromosomes (BACs) have now been used to sequence those genomes in increments. The entire human genome has been fragmented into fragments of 150–200 kb and cloned into BAC vectors ***(42)***. Each BAC is sequenced by the shotgun method and used to form contiguous sequences with overlapping BACs. Physical mapping methods can be used to position members of the BAC library on the intact genome, but this is very costly and a bottleneck to high-throughput sequencing. Instead, sequencing the ends of each BAC clone cre-

ates a database of sequence-tagged sites (STS) that can be used to develop a minimal tiling path ***(43)***. Once a BAC is sequenced, it is the sequence that is used to find the neighboring clone through its STS and not through the physical map. The ultimate application of shotgun DNA sequencing, however, has come recently as a proposal to drop the intermediate step of cloning the human DNA fragments into a BAC vector and to clone them as small fragments directly into sequencing templates ***(44)***. Athough it remains to be seen if an assembly of sequence reads can be made with a 3 billion base pair genome, both approaches to sequence the human genome end up sequencing inserts with universal primers as we did with an 8-kb viral genome project 20 years ago ***(29)***.

Acknowledgments

Most of the initial work on M13 and pUC vectors was supported by the Deutsche Forschungs Gemeinschaft, the Department of Agriculture Grant No. 5901-9-0386, and the Department of Energy and AC02-81ER10901, respectively.

References

1. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **293,** 487–491.
2. Scott, J. and Smith, G. (1990) Searching for peptide ligands with an epitope library. *Science* **249,** 386–390.
3. Edgell, M. H., Hutchison, C. A., III, and Sclair, M. (1972) Specific endonuclease R fragments of bacteriophage ΦX174 deoxyribonucleic acid. *J. Virol*. **9,** 574–582.
4. Beaucage, S. L. and Caruthers, M. H. (1981) Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22,** 1859–1862.
5. Hutchison, C. A., III, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253,** 6551–6560.

6. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., and Hall, B. D. (1979) Sequence of the gene for iso-1-cytochrome C in *Saccharomyces cerevisiae*. *Cell* **16,** 753–761.
7. Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19,** 259–268.
8. Zoller, M. and Smith, M. (1982) Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res.* **10,** 6487–6500.
9. Norrander, J., Kempe, T., and Messing, J. (1983) Improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* **26,** 101–106.
10. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H., and Fischer, H. (1973) Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA. *Proc. Natl. Acad. Sci. USA* **70,** 1209–1213.
11. Hofschneider, P. H. (1963) Untersuchungen über "kleine" *E. coli* K12 Bacteriophagen M12, M13, und M20. *Z. Naturforschg.* **18b,** 203–205.
12. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269,** 496–512.
13. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. V., Heynecker, H. L., Boyer, H. W., Crosa, J. W., and Falkow, S. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2,** 95–113.
14. Vovis, G. F. and Ohsumi, M. (1978) The filamentous phages as transducing particles, in *The Single-Stranded DNA Phages* (Denhardt, D. T., Dressler, D., and Ray, D. S., eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 445–448.
15. Herrmann, R., Neugebauer, K., Zentgraf, H., and Schaller, H. (1978) Transposition of a DNA sequence determining kanamycin resistance into the single-stranded genome of bacteriophage fd. *Mol. Gen. Genet.* **159,** 171.
16. Salivar, W. O., Henry, T. J., and Pratt, D. (1967) Purification and properties of diploid particles of coliphage M13. *Virology* **32,** 41–51.
17. Mandel, M. and Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53,** 159–162.

18. Malamy, M. H., Fiandt, M., and Szybalski, W. (1972) Electron microscopy of polar insertions in the *lac* operon of *Escherichia coli*. *Mol. Gen. Genet*. **119,** 207–222.
19. Landy, A., Olchowski, E., and Ross, W. (1974) Isolation of a functional *lac* regulatory region. *Mol. Gen. Genet.* **133,** 273–281.
20. Messing, J. (1991) Cloning in M13 phage or how to use biology at its best. *Gene*, **100,** 3–12.
21. Messing, J., Gronenborn, B., Müller-Hill, B., and Hofschneider, P. H. (1977) Filamentous coliphage M13 as a cloning vehicle: insertion of a *Hin*dII fragment of the *lac* regulatory region in the M13 replicative form in vitro. *Proc. Natl. Acad. Sci. USA* **74,** 3642–3646.
22. Messing, J. and Gronenborn, B. (1978) The filamentous phage M13 as carrier DNA for operon fusions in vitro, in *The Single-Stranded DNA Phages* (Denhardt, D. T., Dressler, D., and Ray, D. S., eds.) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 449–453.
23. Gronenborn, B. and Messing, J. (1978) Methylation of single-stranded DNA in vitro introduces new restrictions endonuclease cleavage sites. *Nature* **272,** 375–377.
24. Dotto, G. P. and Zinder, N. D. (1984) Reduction of the minimal sequence for initiation of DNA synthesis by qualitative and quantitative changes of an initiator protein. *Nature* **311,** 279–280.
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
26. Heidecker, G., Messing, J., and Gronenborn, B. (1980) A versatile primer for DNA sequencing in the M13mp2 cloning system. *Gene* **10,** 69–73.
27. Messing, J. (1979) A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. *Recombinant DNA Tech. Bull.* NIH Publication No. 79-99, 2, No. 2, 43–48.
28. Messing, J., Crea, R., and Seeburg, P. H. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9,** 309–321.
29. Gardner, R. C., Howarth, A. J., Hahn, P. O., Brown-Leudi, M., Shepherd, R. J., and Messing, J. (1981) The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucleic Acids Res*. **9,** 287l–2888.
30. Holden, C. (1991) Briefings. *Science* **254,** 28.
31. Messing, J. (1983) New M13 vectors for cloning. *Methods Enzymol*. **101,** 20–78.

32. Larson, R. and Messing, J. (1982) Apple II software for M13 shotgun DNA sequencing. *Nucleic Acids Res*. **10,** 39–49.
33. Larson, R. and Messing, J. (1983) Apple II computer software for DNA and protein sequence data. *DNA* **2,** 31–35.
34. Hackett, P. H., Fuchs, J. A., and Messing, J. (1984) An introduction to recombinant DNA techniques, in *Basic Experiments in Gene Manipulation,* Benjamin-Cummings, Menlo Park, CA.
35. Messing, J., Vieira, J., and Gardner, R. (1982) Codon insertion mutagenesis to study functional domains of β-lactamase, in *In Vitro Mutagenesis,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 52.
36. Messing, J. and Seeburg, P. H. (1981) A strategy for high speed DNA sequencing, in *Developmental Biology Using Purified Genes* (Brown, D. and Fox, F., eds.), ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. XXIII. Academic, NY, pp. 659–669.
37. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp and pUC vectors. *Gene* **33,** 103–119.
38. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, CR., Heiner, C., Kent, S. B., and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679.
39. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238,** 336–341.
40. Ewing, B. Hillier, L. Wendl, M. C., and Green, P. (1998) Base-calling of automated sequencer traces using phred - I -accuracy assessment. *Genome Research* **8,** 175–185.
41. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8,** 186–194.
42. Osoegawa, K., Woon, P-Y, Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52,** 1–8.
43. Venter, J. C., Smith, H. O., and Hood, L. (1996) A new strategy for genome sequencing. *Nature* **381,** 364–366.
44. Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998) Shotgun sequencing of the human genome. *Science* **280,** 1540–1542.

3

M13 Sequencing

Qingzhong Yu

1. Introduction

The sequencing of DNA has undergone rapid improvement since the introduction of the chain-termination DNA sequencing method ***(1)*** and the construction of convenient single-stranded DNA cloning vectors, such as the bacteriophage M13 cloning vectors and their derivatives ***(2,3)***. The chain-termination method involves the synthesis of a DNA strand by a DNA polymerase in vitro using a single-stranded DNA template. Synthesis is initiated at only one site where an oligonucleotide primer anneals to the template. The synthesis reaction is terminated by the incorporation of a nucleotide analog that will not support continued DNA elongation. The chain-terminating nucleotide analogs are the 2', 3'-dideoxynucleotide 5'-triphosphates (ddNTPs), which lack the 3'-OH group necessary for DNA chain elongation. When proper mixtures of deoxynucleotide triphosphates (dNTPs) and one of the four ddNTPs are used, enzyme-catalyzed polymerization will be terminated in a fraction of the population of chains at each site where the ddNTP is incorporated. Four separate reactions, each with a different ddNTP, give complete sequence information. A radioactively labeled nucleotide is included in the synthesis, so that the labeled chains of various lengths

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

can be visualized by autoradiography after separation by electrophoresis in a polyacrylamide gel.

As chain termination methods work best when using single-stranded templates, a convenient method of producing single-stranded copies of cloned DNA is essential. The popular M13 cloning vectors and their cousins, the plasmids containing M13 or similar replication origins, solve this problem elegantly ***(2)***. Pure single-stranded DNA can rapidly and easily be prepared from these vectors as described in **ref. *4***. As the popular multiple cloning sites in these vectors are all derived from similar sequences, one primer can be used for the sequencing of insert DNA in most of the related vectors ***(2,3)***. The original chain-termination method uses the large fragment (Klenow enzyme) of *Escherichia coli* DNA polymerase I to synthesize new strands in the sequencing reaction ***(1)***. This enzyme is still used today. Other enzymes such as *Taq* polymerase, T7 polymerase, and Sequenase™ are also widely used. Each enzyme has its own particular properties and qualities, and the choice of polymerase will depend on the type of template and the sequencing strategy employed. In this chapter, the dideoxy chain-termination method using Klenow enzyme and a single-stranded DNA template derived from M13 cloning vectors is described.

2. Materials

1. Template DNA: The single-strand M13 DNA to be sequenced is dissolved in H_2O or TE at a concentration of approx 0.2 µg/µL (*see* **Note 1**).
2. Primer: 1 pmol/µL in H_2O or TE (*see* **Note 2**).
3. TE buffer: 10 m*M* Tris-HCl, pH 8.0, 1 m*M* EDTA, pH 8.0.
4. TM buffer: 100 m*M* Tris-HCl, pH 8.5, 50 m*M* $MgCl_2$.
5. dNTPs: dATP, dCTP, dGTP, and dTTP dissolved in TE to 50 m*M*; dilution of these stocks resulting in 0.5 m*M* working solution. Store at –20°C.
6. ddNTPs: ddATP, ddCTP, ddGTP, and ddTTP dissolved in TE to 10 m*M* as a working solution. Store at –20°C.

7. Four dNTP/ddNTP mixes:
   A-mix: 125 μ*M* dCTP, 125 μ*M* dGTP, 125 μ*M* dTTP, 8 μ*M* ddATP
   C-mix: 6.25 μ*M* dCTP, 125 μ*M* dGTP, 125 μ*M* dTTP, 60 μ*M* ddCTP
   G-mix: 125 μ*M* dCTP, 6.25 μ*M* dGTP, 125 μ*M* dTTP, 100 μ*M* ddGTP
   T-mix: 125 μ*M* dCTP, 125 μ*M* dGTP, 6.25 μ*M* dTTP, 400 μ*M* ddTTP
   These solutions are prepared using the working nucleotide solutions and stored at –20°C (*see* **Notes 3** and **4**).
8. Chase mix: 0.5 m*M* dATP, 0.5 m*M* dCTP, 0.5 m*M* dGTP, 0.5 m*M* dTTP. Store at –20°C.
9. 0.1 *M* dithiothreitol (DTT). Store at –20°C.
10. [^{35}S], α-labeled dATP: Specific activity should be 1000–1500 Ci/mmol, and 10–12.5 mCi/mL.
11. Klenow/label mix (for eight sequencing reactions) 62 μL of 12 m*M* DTT, 3.2 μL of [^{35}S]-dATP (10 μCi/μL), 2.5 μL of Klenow enzyme (5 U/μL, *see* **Note 5**).
12. Stop solution: 100 mL of deionized formamide, 0.1 g of xylene cyanol FF, 0.1 g of bromophenol blue, 2 mL of 0.5 *M* EDTA. Store at –20°C.
13. Microtiter plate: Nonsterile, untreated 96-well U-shaped microtiter plates such as Falcon 3911 or heat-resistant polycarbonate microtiter plates.
14. Gel fixing solution: 10% (v/v) glacial acetic acid, 10% (v/v) methanol.
15. Whatman 3MM paper.
16. Saranwrap®, parafilm, or tape.
17. X-ray film: Kodak BioMax MR or equivalent.

3. Methods

3.1. Annealing Template and Primer

1. In a small microfuge tube, set up the following reaction to a total of 10 μL: 5 μL of template DNA (0.2 μg/μL); 1 μL of primer (1 pmol/μL); 1 μL of TM buffer; 3 μL of H_2O.
2. Incubate the tube in a 55°C waterbath for 30 min (*see* **Note 6**).

3.2. Labeling and Termination Reactions

1. Centrifuge the annealing tube briefly to recover any condensation, and aliquot 2.5 μL of the annealing reaction into four wells, marked A, C, G, or T, of a 96-well U-shaped microtiter plate.

2. Add 2 μL of the respective dNTP/ddNTP mixes into the corresponding wells (e.g., A-mix to A well, C-mix to C well, and so on).
3. Add 2 μL of the Klenow/labeling mix to every well. Mix the reagents by pipeting up and down several times or by brief centrifugation of the plate (covered with a lid or sealed with a piece of parafilm or tape) in a swingout centrifuge. Incubate the plate at room temperature for 20 min (*see* **Note 7**).
4. Add 2 μL of the chase mix to every well and mix the reagents as in **step 3**. Incubate at room temperature for an additional 20 min (*see* **Note 7**).
5. Stop the reaction by adding 2 μL of the stop solution to each well and mix the reagents as in **step 3**.
6. Denature the reaction products by heating the plate on a 80°C heating block or in a 80°C oven for 5 min.
7. Load 2 μL of each sequencing reaction onto a sequencing gel. The remaining materials of the sequencing reactions can be stored at –20°C and used for another sequencing run if necessary. Be sure to denature the sequencing mixture again before loading onto a DNA sequencing gel.

3.3. Autoradiography

1. After finishing the sequencing gel electrophoresis, remove the glass plate from the sequencing equipment.
2. Remove one of the plates so that the gel sticks to one glass plate only. Place the gel on the plate into a horizontal box.
3. Fix the gel by covering it with 2 L of gel fixing solution. Incubate at room temperature for 30 min with occasional agitation to remove urea from the gel (*see* **Note 8**).
4. Transfer the gel to Whatman 3MM paper and cover it with Saranwrap®. Dry the gel on a gel dryer at 80°C for 1–2 h.
5. Expose the gel to a X-ray film for 1–2 d by direct contact between the dried gel and the film emulsion.
6. Develop the film according to the manufacturer's instructions.

4. Notes

1. The yield of single-stranded M13 DNA is approx 5–10 μg from a standard 1.5-mL cell culture preparation. Typically, 1 μg of the M13 DNA template is sufficient for one set of sequencing reactions.

2. Primers of different length (15–26 nt) that anneal to different positions (e. g., –20 and –40 from the cloning site *Pst*I) in the M13 vectors are available from a number of commercial suppliers. There is no detectable difference in the quality of the sequence obtained from single-stranded M13 templates with different length primers. Usually a single "universal" M13 primer can be used to sequence different target DNAs in the M13 vectors.
3. The ratio of dNTP and ddNTP within a particular mix influences the rate of termination within that reaction. Using the dNTP/ddNTP mixes described in this protocol, sequence 250–300 bases from the primer should be readily obtained. If a longer sequence or the first 10–20 nt of the template from the primer is to be sequenced, the ratio of dNTP and ddNTP in the reaction mix can be adjusted. Alternatively, different primers that anneal to the proper position on the template may be used.
4. Sequence compressions occur when the DNA (usually G- or C-rich), synthesized by the DNA polymerase, does not remain fully denatured during gel electrophoresis. Compression artifacts can be eliminated by the use of nucleotide analogs such as dITP or 7-deaza-dGTP to substitute dGTP in the dNTP/ddNTP mix, as the product DNA containing these analogs forms weaker base pairs with dC, and are more readily denatured during gel electrophoresis.
5. Klenow enzyme should be kept constantly cold whenever possible. Klenow/labeling mix should be prepared just before use.
6. Annealing of the template DNA to primer can also be achieved by heating the annealing reaction mix at 65°C for 2 min, then cooling slowly to 30°C over a period of approx 30 min. Once annealed, the template/primer can be stored at –20°C for several weeks without noticeable deterioration.
7. Sequences with strong secondary structure in a template DNA may cause Klenow enzyme to pause during the nascent strand DNA synthesis, resulting in the appearance of bands across all four lanes in the gel. This problem can sometimes be resolved by performing the sequencing reactions at 37°C, or by sequencing a subclone of the template DNA in which the secondary structure is eliminated.
8. The gel fixing solution should be reused, but not more than four times. This will help to reduce the amount of radioisotope waste.

References

1. Sanger, F., Nicklen, S., and Coulson, A. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

2. Messing, J. (1983) New M13 vectors for cloning. *Methods Enzymol*. **101,** 20–79.
3. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33,** 103–119.
4. Tomley, F. (1996) M13 phage growth and single-stranded DNA preparation, in *Methods in Molecular Biology*: *Basic DNA and RNA Protocols*, vol. 58, (Harwood, A. J., eds.), Humana Press, Totowa, NJ, pp. 359–362.

4

Primer Design and Primer-Directed Sequencing

Ulrike Gerischer and Peter Dürre

1. Introduction

The dideoxy chain termination DNA sequencing procedure introduced in 1977 ***(1)*** has the advantage of being fast, simple to perform, and very accurate. Therefore, it became the method of choice to obtain several hundred bases of sequence information per reaction. The procedure is based on the enzymatic elongation of oligonucleotides that are complementary to the single-stranded DNA template. Labeling of the synthesis products is achieved either by incorporation of radioactively or fluorescently labeled nucleotides or by extending primers labeled respectively. Chain extension competes with the infrequent but specific termination by incorporation of a dideoxyribonucleotide. The products of four nucleotide-specific reactions can be separated on a polyacrylamide gel. In the case of a radioactive labeling system, an autoradiogram of such a gel povides the sequence information; fluorescently labeled fragments are used in automated sequencers where a laser beam serves as the detector (*see* Chapters 6, 9–13, and 15); biotinylated fragments can be detected after transfer to a nylon membrane through chemoluminescence ***(2)***. The synthetic oligonucleotide primers required for the synthesis of the labeled strands can easily be synthesized in automatic synthesizers. The service of

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

synthesizing custom-designed oligonucleotides, in many cases with a choice of various modifications, is offered by many companies. For the design of new primers, a few considerations need to be made to obtain efficient sequencing reactions. In addition, so-called "universal sequencing" and "reverse sequencing" primers, complementary to areas flanking the polylinker region in M13 *lac* cloning phages and plasmids, are commercially available.

In principle, there are two different ways of obtaining DNA sequence information by the dideoxynucleotide method. The first strategy involves the creation of a set of subclones to bring different areas of the DNA fragment to be sequenced in proximity of the primer binding site of the vector. This allows for sequencing of the respective fragment with one oligonucleotide and can be achieved using different subcloning strategies. One possibility would be random subcloning after physically breaking up the original fragment (e.g., by shearing). This methodology requires a high number of recombinant clones, many time-consuming and costly sequencing reactions, and a computer for putting the data together. Alternatively, defined subclones can be constructed after restriction mapping the area of DNA to be sequenced, characterized, and finally sequenced. This approach is more straightforward, but also takes a large amount of work and time. As a third possibility, a set of nested deletions can be introduced into the original DNA fragment, originating close to the primer binding site. These nested deletions extend various lengths along the target DNA, thus bringing the area next to the deletion into sequencing range. Again, this procedure requires additional recombinant DNA work on the original clone and the preparation of many DNA templates.

The second approach is based on starting sequencing with primers that are complementary to a known sequence. This can either be the vector or any other sequence adjacent to the DNA to be analyzed. Sequencing proceeds by successive synthesis of new primers at the edges of the newly obtained sequence in such a way that their 3' ends are pointing off into the unknown target DNA ("primer walking method"). This method has the advantage of not requiring the subcloning procedures. The sequencing work can be done on the

original clone. The design of the second primer depends on the result of the first sequencing reaction. For one individual plasmid, only one sequencing reaction can be performed in one direction at a time. Therefore, it is often advisable to produce a small set of subclones and sequence these from both ends simultaneously using the "primer walking method."

In this chapter, we provide information concerning the design of the sequencing primers, as well as a complete outline of a standard seqencing procedure using these primers with plasmid DNA as a template and T7 polymerase ***(3)***.

2. Materials

2.1. Primers

Various primers for a number of commonly used cloning vectors are commercially available (e.g., from Amersham Pharmacia Biotech, Gibco-BRL, MWG Biotech GmbH, New England Biolabs, Sigma, Stratagene). In addition, many universities or other research institutions have their own service facilities for the synthesis of custom-designed primers. For sequencing, oligonucleotides should be deprotected. High performance liquid chromatography (HPLC) purification is not necessary, as the coupling efficiency of the synthesis process lies between 98 and 99.5%. Oligonucleotides are generally delivered in a lyophilized state with information about the amount and instructions for the preparation of a solution of a certain molarity (*see* **Notes 1** and **2**). The concentration of a primer in the sequencing reaction should be approx 1 μ*M* (corresponding to 5.6 μg/mL for a 17-mer). However, even much higher concentrations have been used successfully.

2.2. Template Preparation (Double-Stranded Plasmid DNA)

1. Lysozyme solution: 50 m*M* glucose, 50 m*M* Tris-HCl (pH 8.0), 10 m*M* EDTA (pH 8.0). This solution should be filter-sterilized and stored at 4°C. Prior to use, crystalline lysozyme is added (5 mg/mL) and completely dissolved (*see* **Note 3**).

2. Sodium dodecyl sulfate (SDS) solution: SDS (1%, [w/v]) in NaOH (0.2 *M*). This solution should not be stored longer than 1 wk. Store at room temperature.
3. Sodium acetate solution: 3 *M* sodium acetate, pH 5.2. This solution is prepared by dissolving 246.1 g of water-free CH_3COONa in 800 mL of H_2O, adjusting the pH to 5.2 with glacial acetic acid, and then bringing the volume to 1 L. Store at room temperature after sterilization by autoclaving.
4. Phenol–chloroform solution: Melt crystalline phenol at approx 65°C. Only distilled phenol should be used. Equilibrate the liquid with 0.5 M Tris-base, let phases separate, and remove aqueous phase. Add 50 mM Tris-HCl, pH 8.0, and repeat the last three steps until the pH of the phenol solution is approx 8.0. One volume of chloroform solution (*see* **item 8**) is then mixed with one volume of the equilibrated phenol solution to obtain the final reagent. Store in a brown glass bottle (if possible, under a nitrogen atmosphere) at 4°C up to 1 mo.
5. Ethanol: 96 and 70% (v/v).
6. TE buffer: 10 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA, pH 7.5.
7. RNase solution: 10 m*M* Tris-HCl, pH 7.5, 15 m*M* NaCl, RNase A (10 mg/mL). Dissolve enzyme completely and keep the solution at 100°C for 15 min to inactivate possible DNase contamination. Let cool to room temperature, divide in small aliquots, and store at –20°C.
8. Chloroform solution: Mix chloroform and isoamyl alcohol in a ratio of 24:1.

2.3. Sequencing Reactions

2.3.1. Primer Annealing

1. Template: single- or double-stranded DNA, 0.25 μg/μL.
2. NaOH solution: 2 *M* NaOH.
3. Sodium acetate solution: 3 *M* sodium acetate, pH 4.5. For preparation, *see* **Subheading 2.2., item 3**.
4. Ethanol: 96 and 70% (v/v).
5. Primer solution: 0.8 μ*M* in TE buffer or H_2O.
6. Annealing buffer: 1 *M* Tris-HCl, pH 7.6, 100 m*M* $MgCl_2$, 160 m*M* dithiothreitol (DTT).

2.3.2. Labeling Reaction

1. Sequenase Version 2.0 T7 DNA Polymerase, suitable for sequencing: Amersham Pharmacia Biotech. Be careful to keep the enzyme at

–20°C constantly. The polymerase should be removed from the freezer only for the purpose of taking an aliquot, and this should be done quickly. The aliquot is then diluted with dilution buffer and used for sequencing. The dilution can be kept on ice until the sequencing reactions are set up.

2. Dilution buffer: 20 m*M* Tris-HCl, pH 7.5, 5 m*M* DTT, bovine serum albumin (BSA) (50 μg/mL).
3. Radioactive nucleotide: [α-^{32}P]dATP or [α-^{35}S]dATP, 370 MBq (10 mCi)/mL and 370–463 MBq (10–12.5 mCi)/mL, respectively. Both nucleotides are incorporated equally well into the growing oligodeoxyribonucleotide chain by T7 DNA polymerase. Using ^{35}S, sharper bands are obtained in the autoradiogram and, owing to the avoidance of scattering, longer sequences can be deciphered in one run. The shelf life of this isotope is longer, and the radiation dose of the personnel is reduced ***(4)***. Advantages of ^{32}P are that shorter exposure times are required and that the sequencing gel does not need to be dried before autoradiography.
4. Labeling mix: 1.375 μ*M* dCTP, 1.375 μ*M* dGTP, 1.375 μ*M* dTTP, 333.5 m*M* NaCl. This mix can be stored in aliquots at –20°C.

2.3.3. Termination Reaction

1. Termination mixes: For composition, *see* **Table 1** (*see* **Note 4**).
2. Stop solution: 97.5% (v/v) deionized formamide, analytical grade, 10 m*M* EDTA, pH 7.5, 0.3% (w/v) xylene cyanol FF, 0.3% (w/v) bromophenol blue. This solution should be stored at –20°C (under a nitrogen atmosphere, if possible).

T7 DNA polymerase is also available in a kit, together with most of the required solutions (Amersham Pharmacia Biotech). The composition of these solutions as disclosed by the manufacturer is given in the text and in **Table 1**. Use of this kit saves time and provides so-called "read long sequencing mixes" that allow sequencing of up to 1000 nt per run if a high-resolution sequencing equipment is available.

3. Methods

3.1. Primer Design

Working with custom primers, new oligonucleotides are designed according to the results of the first sequencing reactions with prim-

Table 1
Composition of Termination Mixes

Component	A-Mix μ*M*	C-Mix μ*M*	G-Mix μ*M*	T-Mix μ*M*
dATP	93.5	840	840	840
dCTP	840	93.5	840	840
dGTP	840	840	93.5	840
dTTP	840	840	840	93.5
ddATP	14	—	—	—
ddCTP	—	17	—	—
ddGTP	—	—	14	—
ddTTP	—	—	—	14
Tris-HCl, pH 7.6	40	40	40	40
NaCl	50	50	50	50

ers complementary to the beginning of the polylinker region in all M13 *lac* cloning phages and plasmids. Thus, determined sequences serve to plan the next primer and so on (*see* **Note 5**). As a rule for primer design, the following formula has been reported: primer length = 18 + 1 extra nucleotide for each 2% off of 50% G + C ***(5)***. This is because oligonucleotides with too high an A + T content might not prime satisfactorily with special template DNA strands. Most important is the 3' end of the primer: the last six nucleotides should contain 50% G + C. However, working with an A + T-rich organism (*Clostridium acetobutylicum*), we routinely used 17-mer primers ***(6)*** with G + C contents between two and nine nucleotides and never experienced any problems.

New primer and known sequence should overlap well to avoid not being able to read the starting nucleotides of the new part and to have an internal control of each sequence start. As a rule of thumb the 3' end of the primer should be located approx 30–40 nt from the end of the already known sequence.

Before synthesizing a primer, its sequence should be compared carefully to the whole known sequence of template DNA and vector (preferably by means of a computer program). This will identify regions where undesired hybridizations could take place. Again, of

special importance in this respect are the last ten 3'-terminal bases. An 80% homology in this region might cause major background problems, particularly in G + C-rich templates ***(5)***. If undesired homology can be found, a new primer should be designed. Furthermore, a primer should not contain inverted repeats or sequence repetitions.

There is a fast growing number of sites that can be reached through the Internet, which offer information and software concerning questions of primer design, for example:

Java-based Molecular Biology Work Bench (1/97) (http://www.embl-heidelberg.de/~toldo/JaMBW/Contributors.html) includes multiple links to sequence analysis programs including PCR Primer Design and Oligonucleotide Calculator.

Biological Software (http://pantheon.cis.yale.edu/~huckabyj/biosoft.html) informs about various software for biology including primer design programs.

Virtual Genome Center (6/96) (http://alces.med.umn.edu/VGC.html) offers Oligonucleotide T_m to calculate the melting temperature of an oligonucleotide.

Primer3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) is a primer picking program.

Biodisk Biotechnology Education and Software (http://www.biodisk.com) offers the program Right Primer.

Several commercially available computer programs concerned with DNA also contain programs for primer design: PC Gene (IntelliGenetics), DNA-Strider (Christian Marck, CEA, France), Primer Premier™ 4 (Clontech), PRIMO ***(7)***.

In general, it is not possible to predict the functionality of a new primer. Even oligonucleotides that are designed according to all rules might fail. In such a case, synthesis of a different primer is recommended.

3.2. Template Preparation

Isolation procedures of single-stranded phage DNA and double-stranded plasmid DNA suitable for sequencing are described in other chapters of this volume. Various kits distributed by several molecular biology companies are available for the purification of

plasmid DNA. In most cases, the DNA is purified by passage over anion-exchange resin, resulting in a highly purified product. For most plasmids, we found a modification of the method of Birnboim and Doly *(8)* especially well suited to obtain very satisfying sequencing reactions. This procedure is detailed in the following list:

1. Centrifuge 5 mL of bacterial culture grown overnight at 6000*g* for 5 min at 4°C (*see* **Note 6**).
2. Discard supernatant and suspend the sediment in 150 µL of lysozyme solution.
3. Transfer the preparation to a sterilized microcentrifuge tube and incubate for at least 5 min at room temperature.
4. Add 300 µL of SDS solution and mix gently by hand for a few seconds.
5. Immediately add 225 µL of sodium acetate solution.
6. Mix gently by hand for a few seconds and incubate for 15 min at 0°C.
7. Centrifuge for 5 min at 4°C in a microcentrifuge at maximal speed. Use either a model with a cooling system or place the centrifuge in the coldroom.
8. Transfer supernatant to a new microcentrifuge tube using a microliter pipet with sterilized tip. Be very careful to take only the supernatant.
9. Add 600 µL of isopropanol to precipitate the DNA and vortex a few seconds.
10. Incubate for 5 min at room temperature.
11. Centrifuge for 5 min at room temperature in a microcentrifuge at maximal speed.
12. Remove supernatant completely and rinse sediment with 1 mL of cold ethanol (70%).
13. Let the sediment air-dry (3–5 min) and then suspend it in 100 µL of TE buffer.
14. Add 1 µL of RNase solution and incubate for 15 min at room temperature.
15. Add 260 µL of H_2O and 40 µL of sodium acetate solution.
16. Add 1 vol of phenol–chloroform solution. **Caution:** Wear gloves to avoid any harm to your fingers if this solution is spilled.
17. Vortex-mix for a few seconds and centrifuge for 5 min at room temperature in a microcentrifuge at maximal speed.
18. Transfer the upper phase to a new microcentrifuge tube and add 1 vol of chloroform solution.

19. Mix gently by inverting the closed tubes several times.
20. Centrifuge a few seconds for phase separation and transfer upper phase to a new microcentrifuge tube.
21. Add 2.5 vol of ethanol (96%) to precipitate the DNA, vortex-mix gently, and leave at room temperature for 15 min.
22. Centrifuge for 10 min at room temperature in a microcentrifuge at maximal speed.
23. Wash the sediment with 1 mL of ethanol (70%) and dry it.
24. Suspend the DNA in water.

This procedure usually yields 10–20 μg of plasmid DNA from a 5-mL culture.

3.3. Sequencing Reactions

The procedure described in the following subheading represents a modification of the method of Tabor and Richardson ***(3)***.

3.3.1. Primer Annealing

Whereas single-stranded phage DNA can be used directly as a template, double-stranded plasmid DNA must be denatured first.

1. Pipet 8 μL of double-stranded plasmid DNA (containing 2 μg of nucleic acid) into a sterilized microcentrifuge tube (*see* **Note 7**).
2. Add 2 μL of NaOH and vortex-mix the tube briefly.
3. Centrifuge the tube for a few seconds in a microcentrifuge at maximal speed to concentrate the complete solution at the bottom of the tube.
4. Incubate the tube for 10 min at room temperature.
5. Add 7 μL of sterilized H_2O and 3 μL of sodium acetate solution.
6. Add 60 μL of cold (–20°C) ethanol (96%), mix well, and incubate the cup for 20 min at –70°C.
7. Centrifuge for 10 min at room temperature in a microcentrifuge at maximal speed and discard supernatant.
8. Add 300 μL of cold ethanol (70%) and mix briefly.
9. Centrifuge for 10 min at room temperature in a microcentrifuge at maximal speed.
10. Remove the supernatant carefully and dry the sediment under vacuum (3–5 min).

11. Suspend the pellet in 10 μL of sterilized H_2O.
12. Add 2 μL of annealing buffer and 2 μL of primer solution.
13. Incubate for 20 min at 37°C.
14. Keep the cup at room temperature for at least 10 min.

If the sequencing reaction is not carried out subsequently, the solution should be stored at –20°C until needed.

For single-stranded DNA the following procedure can be used:

1. Pipet 10 μL of template (containing 2 μg of DNA) into a sterilized microcentrifuge tube.
2. Add 2 μL of primer solution and 2 μL of annealing buffer.
3. Vortex-mix the tube briefly.
4. Centrifuge the tube for a few seconds in a microcentrifuge at maximal speed to concentrate the complete solution at the bottom of the tube.
5. Incubate the cup for 10 min at 60°C.
6. Keep the cup at room temperature for at least 10 min.

If the sequencing reaction is not carried out subsequently, the solution should be stored at –20°C until needed.

3.3.2. Labeling Reaction

Because T7 DNA polymerase exhibits a high processivity and a high rate of polymerization, the proper sequencing reaction is divided into two parts. In the first step (labeling reaction), low concentrations of the four deoxyribonucleotides and a relatively low temperature result in the formation of short oligodeoxyribonucleotide chains (approx 20–30 bases). Because a radioactive nucleotide is present in the reaction mixture, all fragments become uniformly labeled, which is important for equal band intensities in the autoradiography step. In the second stage (termination reaction), the four standard reaction mixtures are set up, each of which contain high concentrations of the four deoxyribonucleotides and a single dideoxyribonucleotide. Higher temperature and nonlimiting dNTP concentration then allow a high rate of polymerization that is terminated only by incorporation of a ddNTP.

1. Dilute T7 DNA polymerase with dilution buffer to a concentration of 1.5 U/μL.
2. Add 0.8–1 μL of radioactive dATP ($3.7 \cdot 10^5$ Bq) to the microcentrifuge tube containing annealed primer/template.
3. Add 3 μL of labeling mix.
4. Add 2 μL of diluted T7 DNA polymerase.
5. Mix by pipeting several times up and down in the tube with a microliter pipet and a sterilized tip.
6. Incubate for exactly 5 min at room temperature and then proceed directly with the termination reaction (*see* **step 4**).

3.3.3. Termination Reaction

1. Set up appropriate vials for the termination reaction (microcentrifuge tubes or a microtiter plate) and mark them carefully with "A," "C," "G," and "T."
2. Pipet 2.5 μL of the respective termination mix into the corresponding cup or well. **Steps 1** and **2** should be done in advance (e.g., during the incubation period of the annealing reaction).
3. Warm mixtures to 37°C in a water bath (takes approx 1 min).
4. Pipet 4.5 μL of the labeling mixture into each tube or well. Use a fresh tip for each transfer.
5. Incubate the reaction mixture for exactly 5 min at 37°C.
6. Add 5 μL of stop solution to each tube or well.
7. Denature an aliquot (approx 3 μL) by heating for 3 min at 80°C.
8. Put vials on ice.
9. Load 1.8–2 μL of each mixture onto a sequencing gel. The remaining material of the termination reactions can be stored at –20°C and used for further sequencing runs if necessary. Be sure to denature the aliquots before loading onto a gel.
10. To read as many bases as possible, two samples from each reaction mixture should be run on the same gel, with a period of electrophoresis (approx 2–3 h, until the first colored marker band reaches the end of the gel) between the two loadings.

Starting with prepared template DNA and primers, one round of sequencing including the autoradiography can be performed in 24 h (calculating overnight as exposure time).

4. Notes

1. Oligonucleotide solutions are sensitive toward nucleases. We recommend storing oligonucleotides in the lyophilized state and dissolving them directly before use. The solution should be treated with care, always kept on ice, and stored at –20°C.
2. If the concentration of an oligonucleotide is not given, then both the molar and the mass concentration can be calculated from the absorption of the solution at 260 nm:

$$c\ [\text{pmol}/\mu\text{L}] = A_{260\ \text{nm}} \cdot 100/N$$

$$c\ [\mu\text{g}/\mu\text{L}] = A_{260\ \text{nm}} \cdot 0.033$$

   where N equals the number of nucleotides of the oligonucleotide. This is an estimation based on an average composition of bases with an average molecular weight of 333/nucleotide.
3. In many cases, the plasmid preparation works equally well if the first solution is prepared without lysozyme.
4. To avoid band compressions by G–G or G–C band pairing nucleotide analogs such as 7-deaza-dGTP, 7-deaza-dATP, or dITP can be used. Complete premade mixes with c^7dGTP and c^7dATP are available from Amersham Pharmacia Biotech.
5. We strongly recommend checking the template sequence carefully after every elongation. Ambiguous regions can thus be resequenced at once, which might help to avoid later unnecessary primer synthesis. Direct computer analysis will also enable finding matching stretches if sequencing was started from both ends of the template. If only the end of a newly obtained sequence is read and used for new primer design, this could finally lead to unnecessary sequencing of large parts of the vector.
6. Best sequencing results are obtained after isolating the template DNA from logarithmically growing cells. Therefore, cells should be harvested after growth at 37°C until OD_{600nm} reaches 1.5 or after overnight growth at 30°C.
7. If multiple sequencing reactions with the same double-stranded plasmid are planned, it is recommended that a large amount of plasmid be used in the denaturation reaction. This will save time for future sequencing reactions. The denatured DNA should be dissolved in sterilized H_2O and stored in 10-µL aliquots with the correct amount of DNA for one sequencing reaction at –20˚C.

Acknowledgments

We thank Amersham Pharmacia Biotech for providing the composition of T7 sequencing mixes.

References

1. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
2. Debuire, B., Chabli, A., and Frenoy, N. (1993) Fast, manual, nonradioactive method for DNA sequencing. *Clin. Chem*. **39,** 1682–1685.
3. Tabor, S. and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
4. Van Helden, P. D. (1988) Use of ^{35}S nucleotides for DNA sequencing, in *Methods in Molecular Biology, vol. 4: New Nucleic Acids Techniques* (Walker, J. M., ed.), Humana Press, Totowa, NJ, pp. 81–88.
5. Barnes, W. M. (1987) Sequencing DNA with dideoxyribonucleotides as chain terminators: Hints and strategies for big projects. *Methods Enzymol*. **152,** 538–556.
6. Gerischer, U. and Dürre, P. (1990) Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. *J. Bacteriol*. **172,** 6907–6918.
7. Li, P., Kupfer, K. C., Davies, C. J., Burbee, D., Evans, G. A., and Garner, H. R. (1997) PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing. *Genomics* **40,** 476–485.
8. Birnboim, H. C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*. **7,** 1513–1523.

5

Direct Sequencing of DNA Produced in a Polymerase Chain Reaction

Ulrike Gerischer

1. Introduction

The direct sequencing of an area of DNA using the product of a polymerase chain reaction (PCR) as template is a very powerful technique. It is straightforward and fast and therefore the method of choice in situations such as the analysis of mutations. The reliability of such analysis depends greatly on correct sequencing results. Obtaining the DNA sequence after a PCR amplification and a cloning step contains the risk of sequencing mutations that have been introduced into the DNA by the DNA polymerase upon PCR; the likelihood of such mutations depends on the choice of the enzyme ***(1)***. Using the product of a PCR reaction directly as template in a sequencing reaction circumvents this problem. DNA molecules containing changes introduced upon PCR never amount to a major portion of the total PCR product even if the mutations occur early in the PCR reaction. This can be illustrated by the following calculation: If one works with chromosomal DNA of an organism with a genome size of 4000 kbp (which is in the range of bacterial genomes) and uses as little as 1 ng of this DNA as template in a PCR reaction, this amount contains approx 230,000 copies of the genome. If a muta-

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

tion occurs in one molecule in the first cycle of the temperature program, molecules containing this mutation will make up 1/460,000 part of the entire product. A second application of the direct sequencing of PCR products is the determination of unknown sequences ***(2–4)***.

Many different strategies for the direct sequencing of PCR products have been described. Major differences are whether the double-stranded DNA from the first PCR reaction is sequenced directly or the strand that is supposed to be sequenced is enriched first. The latter can be achieved by digesting the other strand using λ exonuclease ***(5)***, by performing a second PCR with the product of the first reaction as a template using only one of the primers ***(6)***, by hybridization of the other strand with an excess of complementary DNA produced from a cloned fragment ***(7)***, or physically by electrophoresis after heat denaturation and extraction of the desired strand from a gel piece ***(8)***. When sequencing double-stranded DNA directly, many procedures include a heat denaturation of the DNA–primer mix followed by snap cooling at –80°C to improve the hybridization of the primer ***(9)***; in other protocols, the hybridization reactions are performed as described for plasmid DNA ***(10)***.

In all cases, the various procedures include a purification step which serves to separate the PCR product from salt, nucleotides, and PCR primers. The methods for this cleaning include columns (e.g., Microspin HR, Amersham Pharmacia Biotech, Sepharose CL-6B Spin Columns, Boehringer, Wizard PCR Preps Kit, Promega), filter methods (this procedure), or gel extraction by electroelution or binding to glassmilk (e.g., Geneclean, Bio101). The gel extraction has the advantage in being able to separate not only the desired PCR product from lower molecular components of the PCR reaction, but also from other bands that might result from unspecific priming.

The sequencing reactions can be carried out with any system used for routine sequencing using, for example, T7 DNA polymerase or Sequenase in standard sequencing procedures or *Taq* polymerase in cycle sequencing. Addition of detergents or dimethyl sulfoxide (DMSO) has been described to improve the results when sequenc-

ing relatively short DNA fragments (100–400 bp, ***11***). A new method has been described that uses boronated nucleotides to perform the PCR reaction and the sequencing simultaneously ***(12)***.

In this chapter, a method is described that has been used for the sequencing of spontaneous mutations in a previously known stretch of DNA of approx 2 kbp ***(13)***. It includes two PCR reactions, one regular amplification and a second so-called asymmetric PCR favoring the synthesis of one strand. The product is cleaned and sequenced using the kit Sequenase Version 2.0 (Amersham Pharmacia Biotech).

2. Materials

2.1. Isolation of Chromosomal DNA

1. Tris-EDTA solution: 50 m*M* Tris-HCl, pH 8.0, 20 m*M* EDTA.
2. Proteinase K solution: Proteinase K (20 mg/mL).
3. Sodium dodecyl sulfate (SDS) solution: 10% (w/v).
4. Phenol/chloroform solution: Melt crystalline phenol at approx 65°C. Only distilled phenol should be used. Equilibrate the liquid with 0.5 *M* Tris-base, let phases separate, and remove the aqueous phase. Add 50 mM Tris-HCl, pH 8.0, and repeat the last three steps until the pH of the phenol solution is approx 8.0. One volume of chloroform solution (*see* **step 5**) is then mixed with one volume of the equilibrated phenol solution to obtain the final reagent. Store in a brown glass bottle (if possible, under a nitrogen atmosphere) at 4°C up to 1 mo.
5. Chloroform solution: Mix chloroform and isoamyl alcohol in a ratio of 24:1.
6. Ethanol: 96 and 70% (v/v) ethanol.
7. Sodium acetate solution: 3 *M* sodium acetate, pH 5.2. This solution is prepared by dissolving 246.1 g of water-free CH_3COONa in 800 mL of H_2O, adjusting the pH to 5.2 with glacial acetic acid, and then bringing the volume to 1 L. Store at room temperature after sterilization by autoclaving.
8. TE buffer: 10 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA.
9. RNase solution: 10 m*M* Tris-HCl, pH 7.5, 15 mM NaCl, RNase A (10 mg/mL). Dissolve enzyme completely and keep the solution at 100°C for 15 min to inactivate possible DNase contamination. Let cool to room temperature, divide into small aliquots, and store at –20°C.

2.2. Symmetric PCR

1. 10X reaction buffer: 100 m*M* Tris-HCl, pH 8.3, 500 m*M* KCl, 15 m*M* $MgCl_2$, 0.01% (w/v) gelatin (Sigma).
2. dNTP mix: 2.5 m*M* of each of the four deoxynucleotides—dATP, dCTP, dGTP, dTTP in water, pH 7.0.
3. *AmpliTaq*® DNA Polymerase (Perkin-Elmer Cetus, Norwalk, 5 U/μL).
4. Primer: oligonucleotide 20 pmol/μL.
5. Template: chromosomal DNA at 0.1 μg/μL.
6. Mineral oil.

2.3. Asymmetric PCR

The same reagents as for the symmetric PCR are required with the following modification:

1. One of the primers should have a 100-fold lower concentration (0.2 pmol/μL).

2.4. Cleanup of the PCR Reaction Prior to Sequencing

1. A sheet of parafilm (approx 10 × 10 cm), cut from a new roll.
2. Ultrafree-MC Filter Units 30,000 NMWL, regenerated cellulose (Millipore, Japan).
3. TE buffer: 10 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA.

3. Methods

3.1. Isolation of Chromosomal DNA

1. Suspend cells from a 5-mL cell suspension in 0.5 mL of Tris-EDTA solution.
2. Add 50 μL of proteinase K solution and 10 μL of SDS solution.
3. Mix and incubate at 30°C until clear.
4. Add 4 mL of H_2O. (This is done to decrease the viscosity of the solution, which helps in the following extraction.)
5. Extract twice with phenol–chloroform solution.
6. Add 0.1 vol of sodium acetate solution and 2 vol of ethanol. Vortex. Centrifuge for 15 min at 15,000 rpm at 20°C (*see* **Note 1**).
7. Rinse with 70% ethanol, then air-dry.

8. Dissolve in 1 mL of TE buffer containing RNase (10 μg/mL).
9. Determine the DNA concentration by estimation from a gel photo or from a measurement of the absorption at 260 nm.

3.2. Symmetric PCR

1. Design and order a primer pair that will amplify the DNA region to be sequenced (*see* **Note 2**).
2. Dissolve the oligonucleotides in sterile water according to the manufacturer's instructions so that you receive a concentration of 20 pmol/μL (that is 0.11 μg/μL for a 17-mer, *see* **Notes 3** and **4**).
3. Set up the PCR reaction by combining in a reaction tube, which fits into your thermocycler (*see* **Note 5**):
   61.5 μL Water
   10 μL10X Reaction buffer
   8 μL dNTP mix
   0.5 μL *AmpliTaq*® DNA Polymerase
   5 μL Primer 1
   5 μL Primer 2
   10 μL Template
4. Overlay the mixture with 50–100 μL of mineral oil.
5. Run a temperature program that will amplify specifically the desired DNA region (*see* **Note 6**).
6. Run an aliquot (10 μL) of the reaction on an agarose gel to control the success of the amplification. You should see one single band of the expected size (*see* **Note 7**).

3.3. Asymmetric PCR

1. Set up a second PCR reaction equivalent to the one described above with the following modifications:
   a. Use 2 μL of the first PCR reaction (approx 50 ng of DNA) as template.
   b. The primers can be the same primers you used previously or internal primers. One primer should be used in a 100-fold smaller amount, so that the synthesis of one strand is favored. This should be the strand, which serves as the template in the subsequent sequencing reaction (and that is always the strand complementary to the one the sequence of which should be determined).

2. Run an asymmetric PCR reaction using the same temperature program as for the symmetric reaction except the initial melting step at 96°C can be omitted.
3. Run an aliquot of the sample (10 µL) on an agarose gel to make sure you have a product (*see* **Note 7**).

3.4. Cleanup of the PCR Reaction Prior to Sequencing (see Note 8)

1. Separate the asymmetric PCR reaction from the mineral oil using a fresh sheet of parafilm. Put the mixture of oil and aqueous phase on the parafilm. Lift one end of the sheet carefully. A drop consisting of the aqueous phase will run "downhill," whereas the oil will stick to the plastic material. Let the drop move on the sheet of parafilm for approx 10 cm so that it separates from all remaining drops of the oil. Then the drop can be transferred into a reaction tube or directly onto the filter (*see* **step 3**).
2. Pipet the aqueous phase of the asymmetric PCR reaction onto an Ultrafree-MC Filter unit.
3. Bring the volume to approx 400 µL with TE buffer.
4. Centrifuge for 2–4 min at 2000*g* (maximum) in a fixed-angle microcentrifuge until the volume of the sample is 10–50 µL. The centrifugation time may vary depending on the amount of DNA. It is critical at this point not to centrifuge all the liquid through the filter.
5. Remove filtrate to a separate tube.
6. Repeat the wash **steps 3–5** twice to ensure complete separation of the DNA from the components of the PCR reaction.
7. After the last wash step, add a small amount of TE buffer to allow gentle membrane rinsing and then pipet out the DNA solution.

3.5. Sequencing Reactions

1. Use the complete asymmetric PCR product or part of it as template for sequencing, depending on the yield.
2. Denature the DNA as described in the kit Sequenase 2.0 using 2 *M* NaOH. Increase the necessary volumes according to the volume of the DNA solution. After the subsequent precipitation, the denatured DNA is dissolved in the volume specified in the manual.
3. Proceed with sequencing using one of the PCR primers or an internal primer for sequencing (*see* **Note 9**).

4. Notes

1. If you do not have a centrifuge that can be run at 15,000 rpm, the DNA can be recovered at much lower speeds if you extend the centrifugation time. I have routinely precipitated DNA in the described procedure in a centrifugation at 5000 rpm for 30 min.
2. Oligonucleotides should be designed carefully. Chapter 11 (on primer design and primer-directed sequencing) contains more information about what needs to be considered, when designing sequencing primers most of which also applies to PCR primers. The chapter also gives a list of programs that can be used for primer design as well as internet addresses with access to information and software. I have routinely used 17-mers without HPLC purification for the sequencing of bacterial DNA as well as for the PCR reaction.
3. Usually the companies manufacturing the oligonucleotides provide you with information about the product including the amount you receive (usually in the lyophilized state) and instructions how to dissolve it to obtain a certain contentration. If this is not the case, then both the molar and the mass concentration can be calculated from the absorption of the solution at 260 nm:

$$c\ [\text{pmol}/\mu\text{L}] = A_{260\,\text{nm}} \cdot 100/N$$

$$c\ [\mu\text{g}/\mu\text{L}] = A_{260\,\text{nm}} \cdot 0.033$$

   where N equals the number of nucleotides of the oligonucleotide. This is an estimation based on an average composition of bases with an average molecular weight of 333/nucleotide.
4. Oligonucleotide solutions are sensitive toward nucleases. I recommend storing oligonucleotides in the lyophilized state and dissolving them directly before use. The solution should be treated with care, always kept on ice, and stored at –20°C.
5. If more than one PCR reaction is performed at the same time, it is advisable to make a mastermix containing water, buffer, dNTP mix, and enzyme. This increases the accuracy and the identity of the individual samples and saves material. Owing to the viscosity of enzyme solutions (usually supplied in 50% glycerol), it is almost impossible to pipet small volumes (such as 1 µL) exactly. In addition, there is always some solution sticking to the outside of the pipet tip.

6. A great volume of literature is available describing the theory and practice of PCR ***(14)***. In brief, the temperature program should have a denaturation step (94°C, 1 min), an annealing step for 1 min, and an extension step (72°C, 1 min). These three steps should be repeated 25–30×. With chromosomal DNA as the template there should be an initial melting step at 97°C. The annealing temperature depends on the composition and length of the primers and can be calculated via the melting point of a hybrid:

$$T_m = 69.3 + 0.41\% \text{ (G + C)} - 650/\text{length of oligonucleotide in bp}$$

   The annealing temperature should be 3°C below the average melting temperature of both primers.
7. Ideally, after the PCR reaction you should see only the desired product band in the gel. In practice, there are often minor bands from unspecific priming. This can be avoided in some cases after finding the perfect PCR conditions for a specific primer pair. On the other side, it might be worth trying to sequence the DNA with an internal primer. In many cases, sequencing data can be obtained despite the presence of other bands. In contrast, after the asymmetric PCR reaction you may see more than one band even if you have no side products from unspecific priming. This is due to the synthesis of both double-stranded and single-stranded DNA.
8. Several other cleanup systems are available for this purpose, which work equally well, for example, MicroSpin HR (Amersham Pharmacia Biotech), Sepharose CL-6B (Boehringer), or Wizard PCR Preps Kit (Promega).
9. Using an internal primer for the sequencing reaction has the advantage of a higher specificity. If the PCR product to be sequenced contains DNA molecules resulting from unspecific priming, these molecules would have the sequence of the PCR primers at their ends and a different sequence in between. With one of the PCR primers as primer of the sequencing reactions, all these different molecules will contribute sequencing signals. There would be signals at the same position in several lanes and the sequence could not be read. Using a primer that hybridizes to a sequence in between the sequences of the two PCR primers can work only with the desired DNA molecules as template and therefore produces only one sequence.

References

1. Kok, R., D'Argenio, D. A., and Ornston, L. N. (1997) Combining localized PCR mutagenesis and natural transformation in direct analysis of a transcriptional regulator gene, *pobR*. *J. Bacteriol*. **179,** 4270–4276.
2. Pfeifer, G. P, and Riggs, A. D. (1996) Genomic sequencing by ligation-mediated PCR. *Mol. Biotechnol.* **5,** 281–288.
3. Lukyanov, K. A., Matz, M. V., Bogdanova, E. A, Gurskaya, N. G, and Lukyanov, S. A. (1996) Molecule by molecule PCR amplification of complex DNA mixtures for direct sequencing: an approach to in vitro cloning. *Nucleic Acids Res.* **24,** 2194–2195.
4. Capuano, V., Galleron, N., Pujic, P., Sorokin, A., and Ehrlich, S. D. (1996) Organization of the *Bacillus subtilis* 168 chromosome between *kdg* and the attachment site of the SP beta prophage: use of long accurate PCR and yeast artificial chromosomes for sequencing. *Microbiology* **142,** 3005–3015.
5. Takagi, S., Kimura, M., and Katsuki, M. (1993) Direct sequencing of PCR products using unlabeled primers. *BioTechniques* **14,** 218–221.
6. Dowton, M. and Austin, A. D. (1994) A simple method for finding optimal conditions for the direct sequencing of PCR products. *BioTechniques* **16,** 816–817.
7. Gal, S. (1996) Sequencing of double-stranded PCR products. *Mol. Biotechnol*. **5,** 159–164.
8. Iizuka, M., Sugiyama, Y., Iida, S., and Sekiya, T. (1996) Direct sequencing of DNA fragments amplified with single primers by polymerase chain reaction without cloning. *Anal. Biochem*. **241,** 136–139.
9. Green, P. M. and Giannelli, F. (1994) Direct sequencing of PCR-amplified DNA. *Mol. Biotechnol.* **1,** 117–124.
10. DuBose, R. F. and Hartl, D. L. (1990) Rapid purification of PCR products for DNA sequencing using sepharose CL-6B Spin columns. *BioTechniques* **8,** 271–273.
11. Bachmann, B., Lüke, W., and Hunsmann, G. (1990) Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res*. **18,** 1309.
12. Porter, K., W., Briley, J. D., and Shaw, B. R. (1997) Direct PCR sequencing with boronated nucleotides. *Nucleic Acids Res.* **25,** 1611–1617.
13. Gerischer, U. and Ornston, L. N. (1995) Spontaneous mutations in *pcaH,G*, structural genes for protocatechuate 3,4-dioxygenase in *Acinetobacter calcoaceticus*. *J. Bacteriol*. **177,** 1336–1347.
14. Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J.(eds.) (1990) *PCR Protocols. A Guide to Methods and Applications.* Academic Press, San Diego, CA.

6

Solid Phase Fluorescent Sequencing of the *CFTR* Gene

Harry Cuppens and Jean-Jacques Cassiman

1. Introduction

1.1. The CFTR *Gene,* CFTR *Diseases, and* CFTR *Mutations*

Cystic fibrosis (CF) is caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene ***(1–5)***. The *CFTR* gene spans about 250 kb at the genomic level ***(6)***. After transcription, a transcript of about 6.1 kb is obtained containing up to 27 exons ***(2,7)***. Several alternatively spliced transcripts have been found ***(8–13)***, the most important one of which lacks exon 9 sequences ***(8)***. The CFTR protein is a glycosylated transmembrane protein ***(14)*** that functions as a chloride channel ***(15)***. *CFTR* is expressed in epithelial cells of exocrine tissues, such as the lungs, pancreas, sweat glands, and vas deferens ***(1)***. Apart from CF, *CFTR* is also involved in other diseases such as congenital bilateral absence of the vas deferens (CBAVD) ***(16,17)***, allergic bronchopulmonary aspergillosis ***(18)***, and bronchiectasis ***(19)***. More than 850 disease mutations have been described in the *CFTR* gene ***(5)***. Apart from these, more than 120 polymorphisms were also found ***(5)***. The majority of these are point mutations, being missense, nonsense, or

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

splice site mutations, as well as small frameshift mutations (insertions and deletions). Large deletions and insertions are rare.

1.2. Sequencing the CFTR Gene for Diagnostic Purposes

The most frequent mutation, Δ*F508*, contributes about 70% of all mutant *CFTR* genes in the Caucasian population ***(3,20)***. Depending on the ethnic origin, five to ten mutations, such as *1717-1G→A*, *G542X*, *G551D*, *R553X*, *W1282X*, and *N1303K*, reach rather high frequencies ***(4)***. Commercial diagnostic tests for these mutations, either based on the reverse dot-blot principle ***(21)***, allele-specific polymerase chain reaction ***(22)***, or oligonucleotide ligation, are available. About 10–15% of all mutant *CFTR* genes, however, remain undetected in these tests. Other assays, such as sequencing, are required to detect the remainder. Indirect assays such as denaturing gradient gel electrophoresis (DGGE) or single strand conformation polymorphism (SSCP) can be used to pinpoint a region that harbors a mutation, after which the mutation can then be identified by sequencing.

1.3. DNA or RNA as Starting Template for Sequencing?

Either DNA or RNA can be used as starting material for sequencing. Both systems have their advantages and disadvantages.

1.3.1. Availability of Tissue that Expresses the CFTR Gene

DNA can be obtained from the majority of cell types. Blood cells are most accessible and are therefore most frequently used for this purpose. RNA can also be obtained from white blood cells if the gene under investigation is expressed in these cells. If this gene is not expressed in white blood cells, other cells that do express the gene need to be isolated to isolate RNA. In general, the latter turn out to be less accessible when compared with white blood cells. Moreover, the number of cells obtained might be very limited. In such cases, white blood cells might still be used for the isolation of transcripts that were generated through illegitimate transcription

(23). However, one should be very cautious with sequencing data obtained from illegitimate transcripts because aberrant transcripts, which apparently are not caused by a mutation event, can be present. Because *CFTR* is hardly expressed in white blood cells, genomic DNA is preferred as starting material for *CFTR* mutation detection by means of sequencing.

1.3.2. Number of Exons in the Gene

The number of exons, which generally correlates with the size of the gene, needs to be taken into consideration. In general, because of the upper limit size of PCR products (approx 2 kb) that can be obtained in standard PCR reactions and because the size of many introns exceeds 2 kb, all exons require a separate amplification step when genomic DNA is used. Amplification therefore becomes more time consuming when the number of exons is large. When RNA/cDNA is used, several exons can be amplified in a single amplification reaction, thereby reducing the number of PCR reactions. On the other hand, long templates are more prone to problems caused by secondary structure which could result in sequencing artefacts, such as the presence of stops. When DNA is used, however, given the rather small size of exons (on average approx 250 nt) and the number of nucleotides that can be read in a single sequencing reaction (350–600 nt), a considerable fraction of nucleotides read will be derived from intronic regions. In general, one is interested only in the analysis of the coding region and the exon–intron junctions.

1.3.3. RNA Extraction and cDNA Synthesis

Because of the fragility of RNA, precautions need to be taken to prevent its degradation. RNA manipulations and storage are therefore not the method of choice in routine diagnostic settings and are kept to a minimum whenever possible. In addition, the cDNA synthesis reaction makes the protocol more complex and time consuming.

1.3.4. Alternative Splicing

Alternative splicing could hamper the interpretation of sequencing data, or even make it impossible, when RNA/cDNA is used as starting material. Indeed, from the point at which alternative splicing occurs, two overlapping sequences will be found. As long as the proportion of the alternatively spliced form is limited (1–20%), the signal from the latter might not interfere with the interpretation of sequencing data of the full-length transcript. Care therefore needs to be taken in the design of the sequencing strategy, such that sequencing primers do not immediately precede alternatively spliced regions. Alternative splicing will, of course, cause no problems when DNA is used as starting material. Given the presence of several alternatively spliced *CFTR* transcripts, some of which reach rather high proportions ***(8–13)***, RNA/cDNA is not preferred as starting material for routine mutation sequencing.

1.3.5. Mutations that Might Remain Undetected

Depending on the starting material used, particular types of mutations might not be detected (**Table 1**). These shortcomings need to be taken into consideration when a mutation sequencing protocol is developed for a particular gene. Some types of mutations cannot be detected when found in compound heterozygosity with another mutation, when either DNA or RNA is used as starting material, such as a deletion of the complete *CFTR* gene. Indeed, this deletion will be interpreted as homozygous normal because only one allele will be amplified. For the same reason, a deletion of a *CFTR* gene region that harbors one exon, or several exons, will not be detected at the genomic level, while it can be detected at the RNA level. Given the size of the promoter region, promoter mutations might also remain undetected at the genomic level because most genomic sequencing protocols investigate only a limited promoter region just upstream of the first exon. A promoter mutation, regardless of its location in the large promoter region, that results in partial promoter activity might be detected at the RNA level. Some types of mutations (nonsense, splice-site, and frameshift) result in

Table 1
Mutations that Might Remain Undetected Depending on the Starting Material Used for Sequencing

DNA	RNA
Deletion of complete gene	Deletion of complete gene
Deletion of complete exon(s)	
Mutation that results in no promoter activity	Mutation that results in no promoter activity
Mutation that results in partial promoter activity	
	Mutation that results in unstable transcript
Intronic mutation	

unstable transcripts, which may therefore be difficult to detect at the RNA/cDNA level, depending on the degree of instability. In the worst case, mutant *CFTR* transcripts may be even completely degraded and therefore not detectable. At the genomic level, the identification of the latter mutations does not cause any problem. Finally, mutations located deep in intronic regions might remain undetected at the genomic level, as most genomic sequencing protocols of large genes investigate only the coding region and exon/intron junctions. These intronic mutations can be detected at the RNA level. Based on the observation of the existence of a high number of nonsense, frameshift, and splice-site mutations in the *CFTR* gene ***(4,5)***, genomic DNA is preferred as starting material.

1.4. Sequencing Protocol

In a first step, the gene under investigation, or parts thereof, is amplified by PCR. These PCR products are subsequently used as sequencing template. Either a linear sequencing protocol or a cycle sequencing protocol can be used. Linear sequencing is usually performed with T7 DNA polymerase, which generates sequencing fragments that are of uniform intensity, a feature that facilitates the detection of mutations in heterozygous state. Cycle sequencing

facilitates sequencing of DNA that exhibits a high secondary structure content. Moreover, it requires smaller amounts of sequencing template and is less labor intensive. Initially, cycle sequencing assays generated sequencing fragments that were not of uniform intensity and therefore complicated the detection of mutations that were present in the heterozygous state. Cycle sequencing assays that have since then been developed ensure a more even intensity of the different sequencing fragments.

For the linear sequencing protocol, a higher quality of sequencing is obtained when the sequencing template is single-stranded. Single-stranded DNA can be obtained by asymmetric PCR. Here, one of the PCR primers is present at a lower concentration than the second primer. An exponential amplification will be initially obtained. Once the primer that is present at a lower concentration becomes exhausted, a linear amplification is obtained such that single-stranded DNA is generated. Another means for preparing single-stranded template is the use of a PCR reaction in which one of the primers is biotinylated. These biotinylated PCR products are subsequently bound to streptavidin attached to a solid support, such as magnetic beads or a sequencing comb. The latter facilitate the isolation of single-stranded biotinylated DNA after alkali denaturation. The sequencing primer itself can be identical to one of the primers used for amplification; however, nonspecific PCR products can then interfere and decrease the specificity. The use of an internal sequencing primer will therefore increase the specificity and purity of the sequencing reactions.

1.5. Strategy for CFTR *Sequencing*

Based on the criteria described in the preceeding, the following strategy for *CFTR* mutation sequencing is used. Genomic DNA extracted from white blood cells is used as starting material. The different exons of the *CFTR* gene are then amplified by PCR with one of the primers biotinylated. Single-stranded sequencing template is then obtained through binding of biotinylated PCR products to streptavidin-coated magnetic beads, after which the nonbio-

tinylated strand is removed by alkali denaturation. The biotinylated strand is then sequenced with an internal fluorescein isothiocyanate (FITC) labeled sequencing primer by means of a linear sequencing protocol.

2. Materials

1. DNA Thermal Cycler 480 (Applied Biosystems, Foster City, CT).
2. Magnetic Particle Concentrator, type E for tubes of Eppendorf type (Dynal MPC-E) (Dynal, Oslo, Norway).
3. MicroSample™ titerplates (Pharmacia-Biotech, Uppsala, Sweden).
4. Heating plate, accuracy ± 0.1°C (Vel-Merck, Leuven, Belgium).
5. ALF DNA sequencer (Pharmacia-Biotech, Uppsala, Sweden).
6. ALF gel cassette assembly (Pharmacia-Biotech, Uppsala, Sweden).
7. AmpliTaq® (Applied Biosystems, Foster City, CT).
8. Ultrapure deoxynucleotide triphosphates (dNTPs) solution (Pharmacia-Biotech, Uppsala, Sweden).
9. 10X TBE: 0.089 *M* Tris-base, 0.089 *M* boric acid, 0.002 *M* EDTA.
10. Dynabeads M-280 streptavidin (10 mg/mL) (Dynal, Oslo, Norway).
11. 2X Binding and washing (B&W) buffer: 10 m*M* Tris-HCl, pH7.5, 1 m*M* EDTA, 2 *M* NaCl, in ultrapure water.
12. AutoRead (1000) sequencing kit (Pharmacia-Biotech, Uppsala, Sweden).
13. 50% Long Ranger™ gel solution (FMC BioProducts, Rockland, ME).
14. Ultrapure urea (Life Technologies, Paisley, United Kingdom).
15. 10% Ammonium persulfate, freshly prepared (Bio-Rad, Hercules, CA).
16. *N*,*N*,*N*',*N*'-tetramethylethylenediamine (TEMED) (Sigma-Aldrich, Steinheim, Germany).
17. Kimwipes® Lite (Kimberly-Clark).
18. Bind-Silane (Pharmacia-Biotech, Uppsala, Sweden).
19. Bind-Silane stock solution: 150 µL of Bind-Silane/40 mL of absolute ethanol.
20. 10% Acetic acid.
21. Silanization solution, freshly prepared: 80% Bind-Silane stock solution, 2% acetic acid.
22. Gel solution: 7% Long Ranger™ gel solution, 7 *M* urea, 1.2X TBE ALF.
23. 10X TBE ALF: 1 *M* Tris-base, 0.83 *M* boric acid, 10 m*M* EDTA.
24. Multiflex Flat 0.4-mm tips (0.5–200 µL) (Sorenson Bioscience, West Salt Lake City, UT).

3. Methods

3.1. Template Preparation

3.1.1. PCR

Except for the large exon 13, all exons and their exon–intron junctions are individually amplified by PCR. All PCR reactions are performed in 100 μL solutions containing 1X *Taq* DNA polymerase buffer (50 m*M* KCl, 10 m*M* Tris-HCl, pH 8.3, 1.5 m*M* $MgCl_2$, 0.01% gelatine), 200 μ*M* of each dNTP, 200 n*M* oligonucleotide primers (**Table 2**), 2.5 U of Ampli*Taq* and 0.5 μg of genomic DNA (*see* **Notes 1–3**). One of the primers carries a biotin moiety. A negative control of amplification is always included for each primer pair. The amplification conditions are: denaturation for 1 min at 94°C, annealing for 1 min at a temperature optimized for each primer pair (**Table 2**), and extension for 1 min at 68°C, for a total of 30 cycles, with a final extension step of 10 min to fully extend any remaining single-stranded DNA. The amplification is carried out in a DNA Thermal Cycler 480. Ten percent of the amplification products is routinely run on 6% polyacrylamide gel electrophoresis (PAGE) in 1X TBE as a quality control for the different amplifications.

3.1.2. Purification of PCR Products and Preparation of Single-Stranded DNA

1. Take (50*n*) μL (*n* = number of samples to be sequenced) M280 streptavidin Dynabeads with a disposable 1 mL syringe and transfer them to an Eppendorf tube (*see* **Note 4**).
2. Put the Eppendorf tube in a Dynal MPC-E collecting magnetic rack. Within a few seconds the beads will adhere to the wall of the Eppendorf tube, after which the supernatants can be pipeted away.
3. Wash the beads 3× with 1 mL of 1X B&W solution, eliminate the washing solution each time with the Dynal MPC-E collecting magnetic rack (a single wash with 400 μL of 1X B&W is sufficient when only one sample is sequenced).
4. Resuspend the beads in (100*n*) μL of 2X B&W solution.
5. Divide 100 μL aliquots of the resuspended beads in *n* Eppendorf tubes.

Table 2
Oligonucleotides Used for Amplification and Sequencing of the *CFTR* Exons (and Their Exon–Intron Junctions), Length of the PCR Products and Annealing Temperatures for Amplification[a]

Ex1i	P.1i-3b	5'-ACT GCT TAT TCC TTT ACC CC-3'		
	PB.1i-5c	5'-biotin-AC TCG GCT TTT AAC CTG GGC-3'	362 bp	62°C
	PF.1i-3c	5'-FITC-AC GCC CTC CTC TTT CGT G-3'		
Ex2i	P.2i-5b	5'-TCC ATA TGC CAG AAA AGT TG-3'		
	PB.2i-3b	5'-biotin-AG CCA CCA TAC TTG GCT CCT-3'	328 bp	57°C
	PF.2i-5c	5'-FITC-CA AAT CTG TAT GGA GAC C-3'		
Ex3i	P.3i-5*	5'-CTT GGG TTA ATC TCC TTG GA-3'		
	PB.3i-3*	5'-biotin-AT TCA CCA GAT TTC GTA GTC-3'	309 bp	50°C
	PF.3i-5b	5'-FITC-TA AGG GAA ATA GGA CAA C-3'		
Ex4i	P.4i-5*	5'-TCA CAT ATG GTA TGA CCC TC-3'		
	PB.4i-3*	5'-biotin-TT GTA CCA GCT CAC TAC CTA-3'	438 bp	57°C
	PF.4i-5b	5'-FITC-GT GTT GAA ATT CTC AGG G-3'		
Ex5i	P.5i-3b	5'-TTA CTA TTA TCT GAC CCA GG-3'		
	PB.5i-5*	5'-biotin-AT TTC TGC CTA GAT GCT GGG-3'	417 bp	55°C
	PF.5i-3c	5'-FITC-CT CCG CCT TTC CAG TTG T-3'		
Ex6ai	P.6Ai-5*	5'-TTA GTG TGC TCA GAA CCA CG-3'		
	PB.6Ai-3*	5'-biotin-CT ATG CAT AGA GCA GTC CTG-3'	385 bp	57°C
	PF.6Ai-5b	5'-FITC-GT TAG TTT CTA GGG GTG G-3'		
Ex6bi	P.6Ci-3b	5'-CAT TTG ATT GTC ACA AAC ATC-3'		
	PB.6Ci-5*	5'-biotin-TG GAA TGA GTC TGT ACA GCG-3'	444 bp	55°C
	PF.6Ci-3c	5'-FITC-AG GTG GAA GTC TAC CAT G-3'		
Ex7i	P.7i-5b	5'-AGA CCA TGC TCA GAT CTT CC-3'		
	PB.7i-3b	5'-biotin-CA AAG TTC ATT AGA ACT GAT C-3'	409 bp	55°C
	PF.7i-5d	5'-FITC-TG AAA AAT AAA ATA ACA TCC TG-3'		
Ex8i	P.8i-5*	5'-TGA ATC CTA GTG CTT GGC AA-3'		
	PB.8i-3*	5'-biotin-TC GCC ATT AGG ATG AAA TCC-3'	359 bp	50°C
	PF.8i-5b	5'-FITC-GA TGT AGC ACA ATG AGA G-3'		
Ex9i	P.9i-3*	5'-ACA GTG TTG AAT GTG GTG CA-3'		
	PB.9i-5*	5'-biotin-TA ATG GAT CAT GGG CCA TGT-3'	560 bp	65°C
	PF.9i-3c	5'-FITC-GA GAC ATG GAC ACC AAA TTA A-3'		

(continued)

Table 2 (*continued*)

Ex10i	P.10i-5*	5'-GCA GAG TAC CTG AAA CAG GA-3'		
	PB.10i-3*	5'-biotin-CA TTC ACA GTA GCT TAC CCA-3'	491 bp	62°C
	PF.10i-5b	5'-FITC-TT GGA GGC AAG TGA ATC C-3'		
	PF.10i-3b(NB)	5'-FITC-CG ATT GAA TAT GGA GCC-3'		
Ex11i	P.11i-5*	5'-CAA CTG TGG TTA AAG CAA TAG TGT-3'		
	PB.11i-3*	5'-biotin-GC ACA GAT TCT GAG TAA CCA TAA T-3'	425 bp	62°C
	PF.11i-5b	5'-FITC-TT AGA AGG AAG ATG TGC C-3'		
Ex12i	P.12i-5*	5'-GTG AAT CGA TGT GGT GAC CA-3'		
	PB.12i-3*	5'-biotin-CT GGT TTA GCA TGA GGC GGT-3'	426 bp	57°C
	PF.12i-5b	5'-FITC-AA GGC AAA TCA TCT ACA C-3'		
Ex13i1	P.13i-3e	5'-GAA TCT GGT ACT AAG GAC AGC C-3'		
	PB.13i-5f	5'-biotin-TG CTA AAA TAC GAG ACA TAT TGC-3'	530 bp	57°C
	PF.13i-3d	5'-FITC-TC GTA TAG AGT TGA TTG GAT TG-3'		
Ex13i2A	P.13i-3A*	5'-TAC ACC TTA TCC TAA TCC TAT GAT-3'		
	PB.13i-5e	5'-biotin-TC CTG TCT CCT GGA CAG AAA C-3'	577 bp	57°C
	PF.13i-3c	5'-FITC-CA TGC TAC ATA TTG CAT TC-3'		
EX13i2B	P.13i-5e	5'-TCC TGT CTC CTG GAC AGA AAC-3'		
	PB.13i-3A*	5'-biotin-TA CAC CTT ATC CTA ATC CTA TGA T-3'	577 bp	57°C
	PF.13i-5d	5'-FITC-CA ATC TTT TAA ACA GAC TGG AG-3'		
Ex14ai	P.14Ai-5*	5'-AAA AGG TAT GCC ACT GTT AA-3'		
	PB.14Ai-3b	5'-biotin-GT ATA CAT CCC CAA ACT ATC-3'	511 bp	50°C
	PF.14Ai-5c	5'-FITC-AG AAT GAC ATC ATA CAT GGC-3'		
Ex14bi	P.14Bi-3b	5'-CAA TAC ATA CAA ACA TAG TGG-3'		
	PB.14Bi-5*	5'-biotin-GA ACA CCT AGT ACA GCT GCT-3'	294 bp	60°C
	PF.14Bi-3c	5'-FITC-AA TCA ACA GAA ATA AAA CAC-3'		
Ex15i	P.15i-5b	5'-GTG CAT GCT CTT CTA ATG C-3'		
	PB.15i-3*	5'-biotin-AA GGC ACA TGC CTC TGT GCA-3'	485 bp	55°C
	PF.15i-5c	5'-FITC-AG ACT CAA GTT TAG TTC C-3'		
Ex16i	P.16i-5*	5'-CAG AGA AAT TGG TCG TTA CT-3'		
	PB.16i-3*	5'-biotin-AT CTA AAT GTG GGA TTG CCT-3'	570 bp	62°C
	PF.16i-5b	5'-FITC-CT GAA TGC GTC TAC TGT G-3'		
Ex17ai	P.17Ai-5d	5'-TTG AGG TGT TTA AAG TAT GC-3'		
	PB.17Ai-3b	5'-biotin-GT ACA CCA ACT GTG GTA AGA-3'	523 bp	57°C
	PF.17Ai-5c	5'-FITC-GC AAT GTG AAA ATG TTT ACT C-3'		

Ex17bi	P.17Bi-5b	5'-TTC AAA GAA TGG CAC CAG TG-3'		
	PB.17Bi-3*	5'-biotin-AT AAC CTA TAG AAT GCA GCA-3'	463 bp	57°C
	PF.17Bi-5c	5'-FITC-CA GGT ACA AGA TAT TAT G-3'		
Ex18i	P.18i-5*	5'-GTA GAT GCT GTG ATG AAC TG-3'		
	PB.18i-3b	5'-biotin-GT GGC TAT CTA TGA GAA GGA-3'	450 bp	57°C
	PF.18i-5b	5'-FITC-TG CCC TAG GAG AAG TGT G-3'		
Ex19i	P.19i-5b	5'-GCC CGA CAA ATA ACC AAG TG-3'		
	PB.19i-3b	5'-biotin-GC TAA CAC ATT GCT TCA GGC-3'	454 bp	57°C
	PF.19i-5c	5'-FITC-AG CAT CAA ACT AAT TGT GA-3'		
Ex20i	P.20i-5*	5'-GGT CAG GAT TGA AAG TGT GCA-3'		
	PB.20i-3b	5'-biotin-TA TGA GAA AAC TGC ACT GGA-3'	472 bp	57°C
	PF.20i-5b	5'-FITC-GT ACA ATA CTG AAT TAT G-3'		
Ex21i	P.21i-3b	5'-CAA AAG TAC CTG TTG CTC CA-3'		
	PB.21i-5b	5'-biotin-GG TAA GTA CAT GGG TGT TTC-3'	377 bp	57°C
	PF.21i-3c	5'-FITC-AT GTC AGC TAT ATC AGC C-3'		
Ex22i	P.22i-3*	5'-TGT CAC CAT GAA GCA GGC AT-3'		
	PB.22i-5*	5'-biotin-AA ACG CTG AGC CTC ACA AGA-3'	562 bp	55°C
	PF.22i-3b	5'-FITC-AT GAT TCT GTT CCC ACT G-3'		
Ex23i	P.23i-5c	5'-TTA GAG TCT ACC CCA TGG TTG-3'		
	PB.23i-3*	5'-biotin-TA AAG CTG GAT GGC TGT ATG-3'	424 bp	57°C
	PF.23i-5b	5'-FITC-AT AGA CAT ATT ATC AAG G-3'		
Ex24i	P.24i-5*	5'-GGA CAC AGC AGT TAA ATG TG-3'		
	PB.24i-3*	5'-biotin-AC TAT TGC CAG GAA GCC ATT-3'	569 bp	55°C
	PF.24i-5b	5'-FITC-TG TGC CAG TTT CTG TCC C-3'		

[a]Exons are numbered according to the classification of Riordan et al. ***(2)***. Each of the exons 6, 14, and 17 were subsequently found to consist of two exons. These were subsequently divided in exons a and b ***(7)***. The oligonucleotides denoted by P and PB (biotinylated group at the 5' ends) are used for amplification of the respective exons of the *CFTR* gene. The oligonucleotides denoted by PF carry a FITC moiety at their 5' ends and are used as sequencing primers for the respective exons. All primers are used for sequencing of the biotinylated DNA strand except for PF.10i-3b (marked by NB), which is used for sequencing of the nonbiotinylated DNA strand in samples heterozygous for the Δ*F508* or Δ*I507* mutation. The oligonucleotides with an asterisk were taken from Zielenski et al. ***(7)***.

6. Add 50 µL of PCR products to 100 µL of resuspended beads and mix.
7. Incubate for 10 min at room temperature to allow binding of the biotinylated PCR products to the beads (vortex-mix 3×, on a regularly basis during the incubation period).
8. Centrifuge the Eppendorf tubes.
9. Collect the beads with the Dynal MPC-E collecting magnetic rack.
10. Wash 2× with 100 µL of 1X B&W solution with the Dynal MPC-E collecting magnetic rack.
11. Resuspend the beads in 7 µL of H_2O.
12. Add 1 µL of 1 *N* NaOH to the resuspended beads, mix with the pipet.
13. Denature the PCR products for 10 min at room temperature.
14. Collect the beads with the Dynal MPC-E collecting magnetic rack.
15. For sequencing of the nonbiotinylated DNA strand, go to **step a**; for sequencing of the biotinylated DNA strand go to **step b**.
    a. For sequencing of the nonbiotinylated DNA strand:
       Isolate the supernatants.
       Neutralize the supernatants with 1 µL of 1 *N* HCl (*see* **Note 5**).
       Add 2 µL of H_2O.
       Proceed with the sequencing reaction.
    b. For sequencing of the biotinylated DNA strand:
       Wash the beads with 100 µL of 1X B&W solution using the Dynal MPC-E collecting magnetic rack.
       Wash the beads with 100 µL of H_2O using the Dynal MPC-E collecting magnetic rack.
       Resuspend the beads in 11 µL of H_2O.
       Proceed with the sequencing reaction.
16. At this stage, the samples may be stored at –20°C.

3.2. Sequencing Reactions

Sequencing is performed with the AutoRead sequencing kit:

1. Add 2 µL of annealing buffer to the single-stranded DNA solution.
2. Add 4 µL of FITC-labeled sequencing primer (1 pmol/µL), and mix with the pipet.
3. Incubate the sample for 20 min at 37°C in order to allow primer annealing to the DNA template.
4. Cool the sample for at least 10 min to room temperature.

5. Centrifuge the Eppendorf tube in a microfuge.
6. Pipet 2.5 μL of the ddNTP sequencing mixes (A, C, G, and T, respectively) in a MicroSample titerplate and store on ice.
7. Dilute stock T7 DNA polymerase to a concentration of 1.5 U/μL with ice-cold enzyme dilution buffer, mix by gentle pipeting, and store on ice (2 μL will be used for each sequencing reaction).
8. Add 1 μL of extension buffer to the annealed template/primer solution, mix with the pipet, and store on ice.
9. Add 2 μL of T7 DNA polymerase (1.5 U/μL) to the annealed template/primer solution, mix with the pipet, and store on ice.
10. Warm the MicroSample titerplate with the ddNTP sequencing mixes for 1 min on a heating plate at 40°C (to facilitate heat transfer, add a small amount of water on the heating plate in which the MicroSample titerplate will be put).
11. Add 4.5 μL aliquots of the annealed template/primer mix at the heated sequencing mixes (A, C, G, and T), mix with the pipet (use a new tip for each transfer).
12. Incubate the reactions for 5 min on the heating plate (40°C) to allow DNA polymerization.
13. Add 5 μL of STOP solution and mix with the pipet in order to stop the sequencing reactions.
14. Store the MicroSample titerplate on ice until the samples are denatured, or store at –20°C for later use.

3.3. Polyacrylamide Gel Preparation and Electrophoresis

The glass plates are obtained from Pharmacia-Biotech. They should be free of gel remnants, dust, and grease. Gel remnants are scraped off from the glass plate with a plastic ruler under a continuous water flow to prevent scratches on the glass plates. The plates are cleaned with hot tap water and diluted liquid detergent, which does not contain fluorescent elements. After washing, the plates are rinsed extensively in tap water and distilled water, and dried with 100% ethanol. The upper sites of the inner glass plates are silanized (25 μL of silanization solution for each plate) to prevent well damage when the comb is removed after polymerization. Lint-free paper towels should be used to remove any remaining dust particles.

3.3.1. Gel Casting

The gel is prepared according to the protocol of the manufacturer. A brief protocol is given below.

1. The gel cassette is put on a level casting table (*see* **Note 6**).
2. Prepare the gel solution in a beaker that is free of dust and mix on a magnetic stirrer (*see* **Note 7**).
3. Add 40 μL of TEMED and 400 μL of freshly prepared 10% ammonium persulfate to the gel solution and mix.
4. Cast the gel using a 50 mL plastic disposable syringe without a needle. Apply the gel solution in even motions back and forth along the notch in the glass plate, with even pressure on the syringe to ensure that there is always excess gel in the notch. Capillary action will draw the gel front down to fill the whole space between the plates. If there is a tendency of the gel front to become uneven, tap the glass plate lightly with a finger. Also tap the plate at the site of the light coupler if the front reaches the light coupler to prevent bubbles around the light coupler. Leave a small excess of gel solution between the "ears" of the glass plate to ensure that no air will be trapped when the comb is inserted.
5. Allow the gel to polymerize for 45–60 min.
6. To ensure perfect laser transfer through the light coupler, carefully remove any urea crystals that may have formed around and on the light coupler with a lint-free paper towel wetted with distilled water, and wipe subsequently dry with a lint-free paper towel.

3.3.2. Preparation of Samples and Gel Electrophoresis

1. Program the electrophoresis settings: voltage: 1500 V; current: 38 mA; power: 34 W; temperature: 42°C; laser power: 3 mW; sample interval: 2 s; process type: heterozygote low.
2. Install the gel cassette and fill the buffer reservoirs with 0.6X TBE ALF.
3. Rinse the wells using a 50 mL syringe with a canula filled with 0.6X TBE ALF.
4. Heat the MicroSample titerplate at 85–90°C for 3 min, safely floated without its lid in an open water bath, to allow complete DNA denaturation.
5. Place the MicroSample titerplate immediately on ice.
6. Load 4 μL of each sequencing reaction (in the order A, C, G, and T) on the gel (*see* **Notes 8** and **9**).

3.4. Analysis of Results

3.4.1. Heterozygote Criteria

When mutations are found in the homozygous state, a different nucleotide will be found at the mutation site when compared with the wild-type sequence. A different pattern is found for mutations that are found in a heterozygous state. Here, PCR templates are derived from two *CFTR* alleles, except when large deletions are involved. Moreover, an equal amount of PCR products will be derived from each allele when genomic DNA is used. Therefore, two nucleotide peaks will be found at the mutation site. In addition, the intensity of the wild-type peak will be decreased by a factor of 2, and the mutant peak will have an intensity that is only half the intensity of neighboring peaks of the corresponding nucleotide type (**Fig. 1**). In case of a frameshift mutation, all nucleotide positions from the start of the frameshift position could potentially be characterized by the presence of two nucleotide peaks (**Fig. 2**). If these criteria are not met, sequencing artifacts should be considered.

3.4.2. Determination of the Nucleotide Sequence by the ALF Software

The ALF software contains algorithms that assigns nucleotides to the different peaks in the sequencing fluoregram. Four different algorithms are available. One is for nucleotide assignment when a homogeneous template is used for sequencing. If more than one peak is detected at a particular nucleotide position, only the peak with the highest relative intensity is used for nucleotide assignment, while the smaller is treated as background. Three algorithms are available when a heterogeneous template is expected, which therefore allow detection of the presence of two nucleotide peaks at a single nucleotide position so that heterozygous sequences can be detected. These three algorithms differ in the sensitivity by which the heterozygous state is detected. For routine diagnostic purposes, the lowest stringent algorithm is sufficient to detect mutations in heterozygote state. When the more stringent algorithms are used,

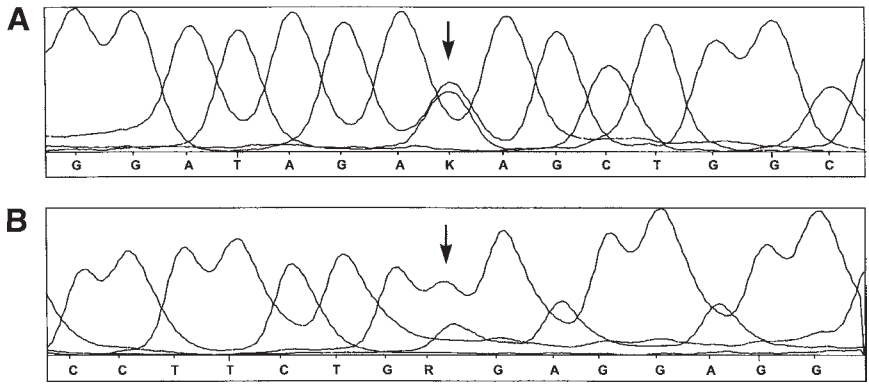


Fig. 1. Examples of missense mutations found in the heterozygous state. **(A)** A "G" to "T" mutation at nucleotide position 310 of the *CFTR* gene in the E60X nonsense mutant. **(B)** A "G" to "A" nucleotide change at position 1335 of the *CFTR* gene in the W401X mutant. The positions of the mutations are indicated by an arrow.

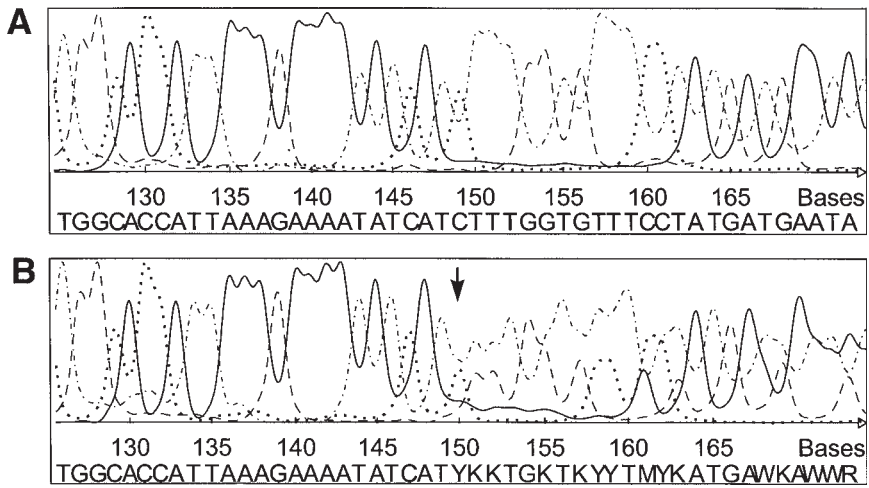


Fig. 2. Example of a deletion mutation. **(A)** Part of the wild-type sequence of exon 10. **(B)** Part of the corresponding exon 10 sequence derived from an individual who is heterozygous for the ΔF508 deletion. The start of the deletion is indicated by an arrow.

more ambiguities will be found because of background signals that are interpreted as mutations. The majority of the mutations can be easily detected by the ALF software (*see* **Figs. 1** and **2**).

3.4.3. Types of Mutations that Might Remain Undetected by the ALF Software

The resolution by which the different peaks are separated decreases along the sequence. **Figure 3** illustrates two identical time intervals of a single sequencing reaction, one starting from nucleotide position 23 and the other from nucleotide position 234. In low-resolution regions, a mutation will be more difficult to detect, especially if the preceding or next nucleotide is identical to one of the nucleotides found in the heterozygous state. It might be even more problematic if both the preceding and next nucleotide correspond to the nucleotides that are found in the heterozygous state. In very rare instances, such a mutation might even remain undetected by the ALF software (*see* **Figs. 4** and **5**). **Figure 4A** shows a wild-type sequence of exon 10, while the two other figures show examples of *E528E* alleles (raw data sequences are shown). The *E528E* is a silent mutation that involves the last nucleotide of exon 10. In **Fig. 4B**, the mutation was detected by the ALF software, while the mutation in **Fig. 4C** was not detected by either one of the three heterozygote ALF algorithms. The processed sequence data of these samples are shown in **Fig. 5**. Strikingly, in comparison with the mutation shown in **Figs. 4B** and **5B**, the mutation in **Figs. 4C** and **5C** is more clear when visually inspected.

It should be noted that these artefacts are rare but can occur. Sequencing of both strands practically eliminates the occurrence of such false-negatives. The latter, however, doubles the sequencing efforts and sequencing time. Visual inspection of the nucleotide assignment is less time consuming and is therefore preferred.

Because of these difficulties of heterozygous mutation detection in low-resolution regions, even when 500–600 nt can be read, only up to the 450 first nucleotides can be routinely interpreted for heterozygous mutation detection.

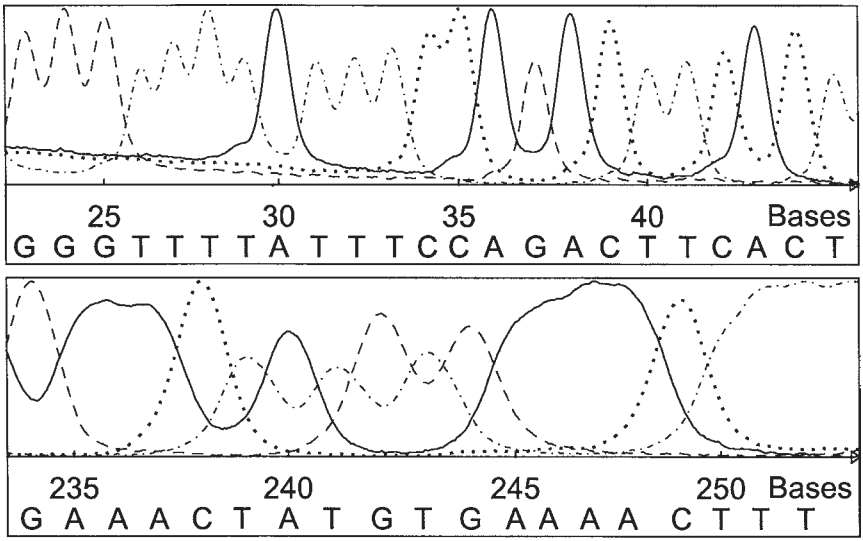

25 30 35 40 Bases
G G G T T T T A T T T C C A G A C T T C A C T
235 240 245 250 Bases
G A A A C T A T G T G A A A A C T T T

Fig. 3

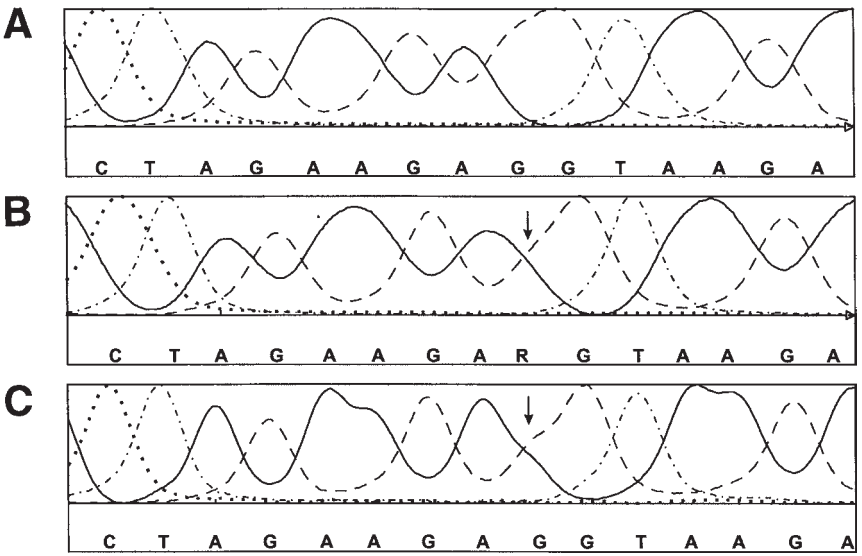

A
C T A G A A G A G G T A A G A
B
C T A G A A G A R G T A A G A
C
C T A G A A G A G G T A A G A

Fig. 4

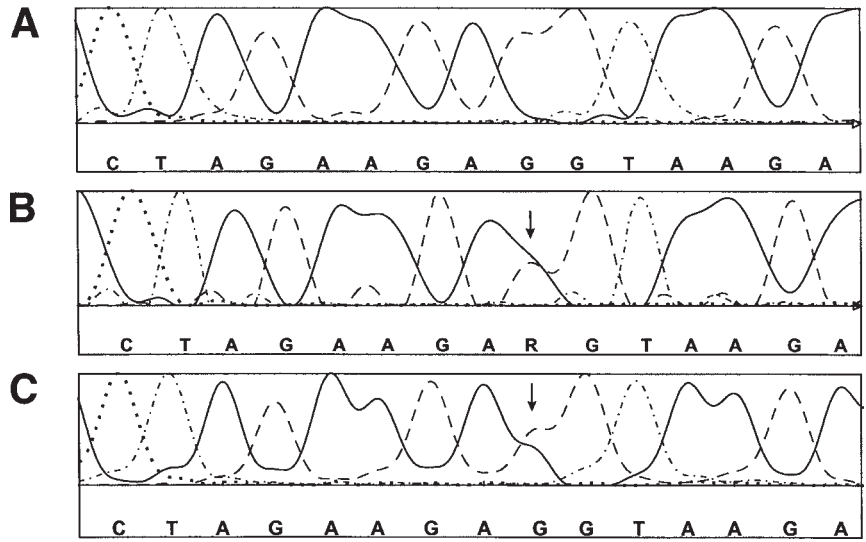


Fig. 5. Illustration of the processed sequencing data of a mutation that remains undetected by the ALF software. Same as in **Fig. 4,** but the processed sequences are shown.

3.4.4. False-Positive Mutations Because of Compressions

If both heterozygote criteria are not met, sequencing artefacts should be considered. An example is shown in **Fig. 6**. **Figure 6A**

Fig. 3. *(opposite page)* Illustration of the decrease in resolution along the sequence. Two identical time intervals of a single sequencing sample are shown. The *upper panel* shows the sequence starting from base position 23; the l*ower panel* shows the sequence data starting from base position 234.

Fig. 4. *(opposite page)* Illustration of the raw sequencing data of a mutation that remains undetected by the ALF software. The mutations sites are indicated by an arrow. **(A)** Part of the wild-type sequence of exon 10. **(B)** The corresponding exon 10 sequence from an individual who is heterozygous for a "G" to "A" change at nucleotide position 1716 and that is detected by the ALF software. **(C)** The corresponding exon 10 sequence from another individual who is heterozygous for the same mutation but that is not detected by the ALF software.

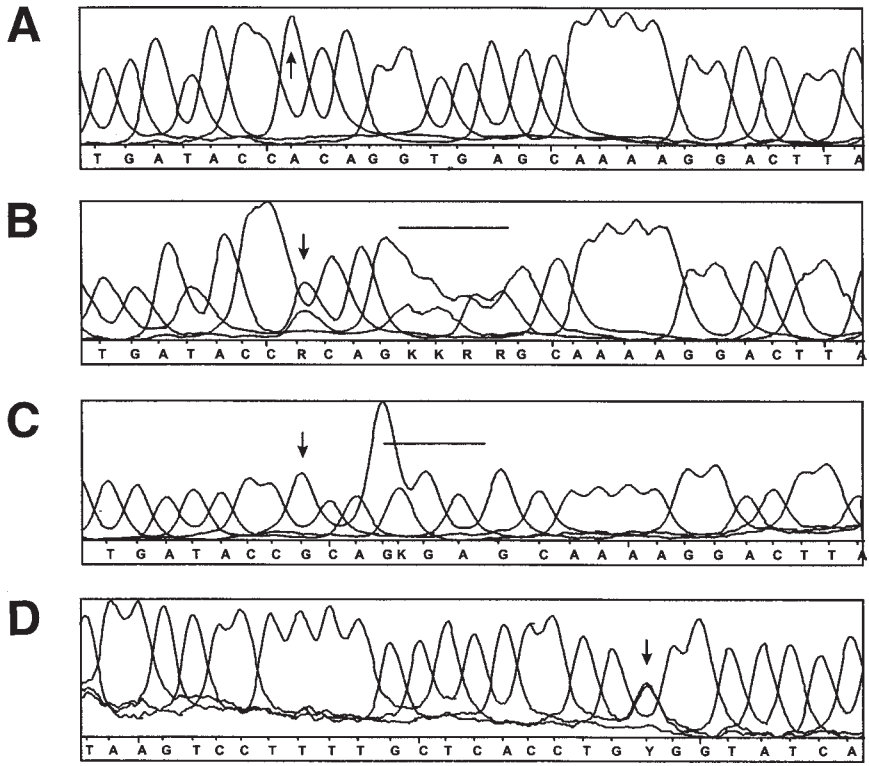


Fig. 6. Illustration of a false-positive mutation because of the presence of a compression. The mutation sites are indicated by an arrow; the location of compressions are indicated by a bar. **(A)** Part of the wild-type sequence of exon 20 and its exon–intron junction. **(B)** The corresponding sequence as in A from an individual who is heterozygous for a "A" to "G" change at nucleotide position 4002 that results in the silent *P1290P* mutation. A complex sequencing pattern is observed that starts four nucleotides downstream the mutation at nucleotide position 4002. **(C)** The corresponding sequence as in A from another individual who is homozygous for the "A" to "G" change at nucleotide position 4002, and who harbors a compression, starting four nucleotides downstream of this mutation. **(D)** The complementary sequence of the sample that is presented in B, obtained by sequencing of the complementary strand.

shows the wild-type sequence of exon 20, and its exon-intron junction. In **Fig. 6B**, a sample is shown that is heterozygous for a silent mutation at codon 1290 (*P1290P*). Moreover, a non-wild-type pat-

tern is observed at four nucleotide positions, starting four nucleotides downstream of the *P1290P* mutation. However, the heterozygote criteria are not always obeyed. In **Fig. 6C**, a sample is shown that carries *P1290P* in the homozygous state. Here, a compression is observed that starts four nucleotides downstream *P1290P*. The aberrant sequencing pattern that starts four nucleotides downstream of *P1290P* in **Fig. 6B** therefore results from compressed nucleotides derived from the sequence termination fragments of the *P1290P* allele, and noncompressed sequence termination fragments of the wild-type allele, rather than because of the presence of a mutation. Of all samples sequenced so far, this compression was seen only in the presence of *P1290P* and never in wild-type sequences. Therefore, the compression is very likely caused by the *P1290P* mutation. The compression was not seen when the complementary strand is sequenced (*see* **Fig. 6D**).

It should again be noted that these artefacts are rare but can occur. Sequencing of both strands practically allows one to prevent such false-positive mutations.

3.4.5. Sequence Alignment and Comparison Software

The mutation is further characterized once it is detected by the ALF software: that is, the nucleotide number of the mutated nucleotide, the amino acid number that is affected, and the amino acid codon change or disruption/creation of splice site are determined. Therefore, the nucleotide sequences are exported to file formats that can then be further analyzed. For this purpose, several software programs for alignment and comparison are commercially or publicly available on the Internet. Recently, the ALF mutation analyzer software (Pharmacia-Biotech, Uppsala, Sweden) has been released that combines all these software applications for easy mutation characterization.

4. Notes

1. The amplification of the different exons is performed in a final volume of 100 μL, such that a second sequencing reaction can be performed in case of sequencing failure. Sequencing is always performed with 50 μL of a PCR reaction solution. Since 10 μL is used for PAGE

analysis, only 40 μL PCR reaction solution can be used as starting template for a second sequencing reaction which still will provide high-quality sequence data. No mineral oil is needed when the amplification is performed in a volume of 100 μL, while mineral oil is required when the amplification is performed in a volume of 50 μL.

2. The concentration of PCR primers used during amplification should not exceed 200 n*M*. Higher concentrations will result in a higher proportion of primers that remain not extended. These will compete with the extended primers for binding to the streptavidin coated magnetic beads. A lower amount of template PCR products will then be isolated, which in turn results in a lower intensity of the sequencing signals.
3. Biotinylated PCR products are stored at –20°C to ensure their stability and quality.
4. When isolating M-280 streptavidin Dynabeads for the preparation of single-stranded DNA, only up to 750 μL of beads should be pipeted per Eppendorf tube, as the beads will be finally resuspended in a double volume of B&W buffer (i.e., 1500 μL).
5. The nonbiotinylated strand can also be used for sequencing; however, the quality of the sequencing reactions is poor compared with the sequencing reactions obtained when the biotinylated strand is used as template. The additional amount of NaCl, because of the denaturation of the double-stranded PCR products and the subsequent neutralization of the solution that contains the nonbiotinylated DNA strand, is most probably responsible for the poorer quality. Moreover, accurately titered 1 *N* NaOH and 1 *N* HCl solutions for alkali denaturation and neutralization are of the utmost importance to obtain the highest quality sequence.
6. When silanizing the upper sites of the glass plates, be careful to silanize only the upper sites of the inner glass plates and not the complete glass surfaces, as this may hamper the separation of the glass plates when cleaned after use. All subsequent cleaning steps should therefore be performed with paper towels, moving them from the lower part to the upper part of the glass plates, and they are always discarded once they came in contact with silanized regions.
7. A 7% Long Ranger sequencing gel is used when screening for heterozygous mutations, instead of the recommended 5% gel. To detect mutations in the heterozygous state efficiently, the resolution of separation of the different nucleotides is of utmost importance, especially when stretches of identical nucleotides are found, rather than the number of nucleotides read. Of the four different nucleotides,

cytosine stretches have the lowest resolution. A better resolution of cytosine stretches is obtained in 7% Long Ranger sequencing gels.

8. If different exons are sequenced, the different sequencing reactions are loaded in decreasing order of length. Here, the length of the longest sequence fragment obtained in a particular sequencing reaction is taken into consideration, not the length of the complete PCR product. Indeed, a large fraction of the sequencing primers is completely extended along the PCR template, that is, no termination nucleotides are incorporated. The sequencing fragment that ends at the last base of the template will therefore be found at a much higher proportion than the other sequencing termination fragments. This fragment might then, because of signal overflow, be detected by detectors of neighboring clones (especially the T lane of the preceding clone and the A lane of the next clone). This might in turn interfere with the interpretation of the sequence data of these neighboring clones, generating two peaks at a particular nucleotide position that could be interpreted as a mutation in heterozygous state, or the presence of a "stop" if signal overflow occurs to more than one detector of the neighboring clones. Although the heterozygote criteria are not met, because the wild-type nucleotide fragment will not be found at a decreased intensity, a heterozygous mutation might be interpreted by a nonexperienced eye. If different exons are sequenced, the different sequencing reactions are therefore loaded in decreasing order of length. In general, when two neighboring clones harbor sequencing reactions of different exons, the last sequencing fragment in the second clone will cause a signal overflow in the preceding clone at a position that coincides with a less important intronic nucleotide. The signal overflow from the first to the second clone may also coincide with a less important intronic nucleotide or may even occur at an elution time that bypasses the elution time of the longest termination fragment in the second clone. The sequencing reactions are loaded in decreasing order of length such that the longer ones are located close to the laser source, while the shorter are located more distant from the laser source. Indeed, the quality of the laser beam in the gel and/or detector sensitivity may be lower at more distant regions from the laser source, for example, because of the presence of dust particles in the gel such that laser signal is absorbed and becomes weaker, or because of a nonhorizontal laser beam.
9. For loading of the sequencing gel, a single multiflex flat tip can be used for loading of all sequencing reactions. After loading of a well,

the tip is washed by pipeting it up and down in the upper buffer solution after which the tip is dried with a paper towel.

References

1. Welsh, M. J., Tsui, L.-C., Boat, T. F., and Beaudet, A. L. (1995) Cystic fibrosis, in *The Metabolic and Molecular Bases of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds.), McGraw-Hill, New York, pp. 3799–3876.
2. Riordan, J. R., Rommens, J. M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., et al. (1989) Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA. *Science* **245,** 1066–1073.
3. Kerem, B.-S., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M., and Tsui, L.-C. (1989) Identification of the cystic fibrosis gene: Genetic analysis. *Science* **245,** 1073–1080.
4. The Cystic Fibrosis Genetic Analysis Consortium. (1994) Population variation of common cystic fibrosis mutations. *Human Mutat.* **4,** 167–177.
5. The Cystic Fibrosis Genetic Analysis Consortium. Cystic Fibrosis Mutation Data Base. http://www.genet.sickkids.on.ca/cftr/
6. Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., et al. (1989) Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science* **245,** 1059–1065.
7. Zielenski, J., Rozmahel, R., Bozon, D., Kerem, B.-S., Grzelczak, Z., Riordan, J. R., Rommens, J., and Tsui, L.-C. (1991) Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. *Genomics* **10**, 214–228.
8. Chu, C.-S., Trapnell, B. C., Murtagh, J. J., Jr., Moss, J., Dalemans, W., Jallat, S., et al. (1991) Variable deletion of exon 9 coding sequences in cystic fibrosis transmembrane conductance regulator gene mRNA transcripts in normal bronchial epithelium. *EMBO J.* **10,** 1355–1363.
9. Slomski, R., Schloesser, M., Berg, L.-P., Wagner, M., Kakkar, V. V., Cooper, D. N., and Reiss, J. (1992) Omission of exon 12 in cystic fibrosis transmembrane conductance regulator (*CFTR*) gene transcripts. *Hum. Genet.* **89,** 615–619.
10. Bremer, S., Hoof, T., Wilke, M., Busche, R., Scholte, B., Riordan, J. R., Maass, G., and Tümmler, B. (1992) Quantitative expression pat-

terns of multidrug-resistance P- glycoprotein (MDR1) and differentially spliced cystic-fibrosis transmembrane-conductance regulator mRNA transcripts in human epithelia. *Eur. J. Biochem.* **206,** 137–149.

11. Will, K., Stuhrmann, M., Dean, M., and Schmidtke, J. (1993) Alternative splicing in the first nucleotide binding fold of CFTR. *Hum. Mol. Genet.* **2,** 231–235.
12. Yoshimura, K., Chu, C.-S., and Crystal, R. G. (1993) Alternative splicing of intron 23 of the human cystic fibrosis transmembrane conductance regulator gene resulting in a novel exon and transcript coding for a shortened intracytoplasmic C terminus. *J. Biol. Chem.* **268,** 686–690.
13. Morales, M. M., Carroll, T. P., Morita, T., Schwiebert, E. M., Devuyst, O., Wilson, P. D., et al. (1996) Both the wild type and a functional isoform of CFTR are expressed in kidney. *Am. J. Physiol.* **270,** F1038–F1048.
14. Gregory, R. J., Cheng, S. H., Rich, D. P., Marshall, J., Paul, S., Hehir, K., et al. (1990) Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature* **347,** 382–386.
15. Bear, C. E., Li, C., Kartner, N., Bridges, R. J., Jensen, T. J., Ramjeesingh, M., and Riordan, J. R. (1992) Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (*CFTR*). *Cell* **68,** 809–818.
16. Chillón, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., et al. (1995) Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N. Engl. J. Med.* **332,** 1475–1480.
17. Zielenski, J., Patrizio, P., Corey, M., Handelin, B., Markiewicz, D., Asch, R., and Tsui, L.-C. (1995) CFTR gene variant for patients with congenital absence of vas deferens. *Am. J. Hum. Genet.* **57,** 958–960.
18. Miller, P. W., Hamosh, A., Macek, M., Jr., Greenberger, P. A., MacLean, J., Walden, S. M., et al. (1996) Cystic fibrosis transmembrane conductance regulator (*CFTR*) gene mutations in allergic bronchopulmonary aspergillosis. *Am. J. Hum. Genet.* **59,** 45–51.
19. Pignatti, P. F., Bombieri, C., Marigo, C., Benetazzo, M., and Luisetti, M. (1995) Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. *Hum. Mol. Genet.* **4,** 635–639.
20. The Cystic Fibrosis Genetic Analysis Consortium. (1990) Worldwide survey of the ΔF508 mutation—report from the cystic fibrosis genetic analysis consortium. *Am. J. Hum. Genet.* **47,** 354–359.
21. Cuppens, H., Buyse, I., Baens, M., Marynen, P., and Cassiman, J.-J. (1992) Simultaneous screening for 11 mutations in the cystic fibrosis

transmembrane conductance regulator gene by multiplex amplification and reverse dot-blot. *Mol. Cell. Probes* **6,** 33–39.

22. Ferrie, R. M., Schwarz, M. J., Robertson, N. H., Vaudin, S., Super, M., Malone, G., and Little, S. (1992) Development, multiplexing, and application of ARMS tests for common mutations in the *CFTR* gene. *Am. J. Hum. Genet.* **51,** 251–262.
23. Chelly, J., Concordet, J. P., Kaplan, J. C., and Kahn, A. (1989) Illegitimate transcription: Transcription of any gene in any cell type. *Proc. Natl. Acad. Sci. USA* **86,** 2617–2621.

7

Shotgun DNA Sequencing

Alan T. Bankier

1. Introduction

When the Human Genome Mapping Project began in earnest, the allocation of large funding and the opening of dedicated research facilities promised much for the rapid improvement of DNA sequencing methods and, as a consequence, sequencing rates and throughput. Today, several years later, some novel approaches to the problem are being developed, such as sequencing by hybridization, and sequence detection using mass spectrometry and biochips, but these are far from in general use.

The greatest improvements have arisen from modifications to sequencing chemistries, for example, by using modified thermostable polymerases, and by improvements to the accuracy and capacity of automated fluorescent sequencers. The combination of these improvements has extended the average read length, greatly increased the number of samples per gel (or number of capillaries per instrument), and shortened run times.

Although average read lengths have risen significantly, limitations still apply to the size of target DNA that can be sequenced in a single run. Unless the sequence to be determined is less than or approx 1 kb, primer walking or subcloning strategies need to be used ***(1)***.

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

If the DNA template is available in sufficient quantities, primer walking is very efficient (*see* Chapter 4). Each round of sequencing, to extend the region of known data, is preceded by sequence evaluation, primer design, and oligonucleotide primer synthesis. If the target is only a few kilobases in length, this poses few problems and sequence can be determined progressively from both ends, at a modest but acceptable rate. For longer targets, and recombinants from cosmids BACs or PACs, this process is prohibitively slow and expensive.

Where very little template is obtainable or the sequence to be determined is large, subcloning must precede sequencing. This presents the possibility of employing walking strategies based upon ordered deletions ***(1–3)***. These methods bring consecutive parts of the DNA to within the range of a single universal primer. The utility of the method, however, is severely limited by the size of the insert that can be sequenced (as the target needs to be cloned in its entirety) and by the considerable time and effort required to generate the deletion library.

When large segments of DNA are to be sequenced, the preferred method is normally random (shotgun) subcloning and sequencing ***(4,5)***. Indeed, this is currently the method of choice for most of the large-scale genomic sequencing projects. The DNA to be sequenced is randomly broken down to smaller fragments, which are cloned into a suitable vector for sequence template preparation. Several methods of producing random breaks are used including disruption in a French press, passing the DNA solution through a fine-gauge needle or bore constriction ***(6)***, enzymatically by using DNase I ***(7)***, and shearing by sonication ***(8)***. The latter three techniques are the most widely used.

Recombinants from the subclone library are then selected at random and sequenced. The relatively short individual sequence reads are compared with each other and overlaps are accumulated to reconstruct the original fragment. Clearly, this approach leads to significant redundancy toward the end, as each recombinant can arise from anywhere within the original DNA fragment. The more sequence data determined, the more redundancy that can be expected.

Assuming a purely random distribution of each subfragment, the level of redundancy can be modeled by the equation:

$$S = L(1 - (1 - r/L)n)$$

where S is the proportion of the sequence determined, L is the total fragment length, r is the average read length, and n is the number of clones sequenced.

The rate of data capture is, not surprisingly, heavily weighted by the length of the individual reads.

This simplistic model assumes that subfragments are produced in a truly random manner (the starting material is circular and that there is no sequence bias in the generation of subfragments). The first of these criteria can be satisfied by using intact recombinant clone material, such as that derived from cosmids, where the vector sequence is so small as to contribute little to the reconstructed sequence and the extra effort required to process it is minimal. An alternative is to self-ligate the isolated insert prior to shearing. Although sequences that are A–T-rich seem more resistant to shearing by sonication, no significant bias is caused by this.

Other key features to note are:

1. It is of benefit to ensure that the subclones are at least as large as the maximum read lengths obtainable; otherwise, much time is wasted sequencing vector ends.
2. If subfragments are shorter than the length of the maximum read length, it is possible to read sequence across the junction of dimer clones, where two or more fragments are cloned together. This will become apparent as data has accumulated, but it can cause considerable confusion until several sequences across the region are gained.
3. If subclones are generated from fragments twice the maximum read length, then the amount of information per recombinant can be doubled by performing forward and reverse reads from each clone.
4. If even larger subfragments are chosen, then each clone provides not only two sequence reads but also additional positional information that links together two sequences, separated by a predictable distance. This can be invaluable at later stages when gaps need to be closed.

Clearly, the larger the fragments are, the better. This is only limited by the vector–host system used: how large a fragment the vector can hold and how stable large constructs are.

The most popular sequencing vector system has traditionally been the M13 series of vectors (*see* Chapters 2 and 3). These have been chosen because template DNA can be produced easily and in large quantities. In addition, the resultant single-stranded template gives much cleaner sequence reads. Unfortunately, this vector system can routinely hold inserts of only up to a few kilobases and stability is a severe problem with some DNA sequences, particularly those containing repetitive or palindromic regions.

Recent improvements in sequencing chemistry, for example, using cloned, modified thermostable polymerases and cycle sequencing using labeled terminators, has greatly improved the quality and read lengths gained from double-stranded templates, and so the availability of suitable vectors is less of a problem.

One of the major benefits of shotgun sequencing is its universal applicability. The technique works independently of fragment length, at a wide range of DNA concentrations, and is only moderately affected by base composition. This means that little intellectual or experimental input is required and a mass-production system can be adopted where any DNA from any source can be put into the processing production line. This is of great benefit to genome centers where uniformity of production means the highest throughput.

There is little point, however, in blindly sequencing random recombinants when redundancy means that toward the end, carrying out many sequence reactions yields only a few new bases. This problem of redundancy is even worse for larger projects. At some point, depending on how large the target is, how long the inserts are, how much time is available, and so on, it is likely that alternative more directed strategies are going to prove more productive in finishing the project. This can take the form of "fishing" the recombinant library for clones that close the gaps, specific subfragment cloning based upon the ends of known contigs or, if you are lucky, simply primer walking on larger clones that may extend into the void.

2. Materials

2.1. DNA Self-Ligation

1. $TE_{0.1}$: 10 m*M* Tris-HCl, pH 8.0, 0.1 m*M* Na_2EDTA.
2. T4 DNA ligase: New England Biolabs.
3. 10X Ligase buffer: 500 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 100 m*M* dithiothreitol (DTT): This buffer is normally supplied with the enzyme.
4. 10 m*M* ATP: This is normally supplied with the enzyme.
5. Phenol equilibrated to pH 8.0 with Tris buffer: Amersham Pharmacia Biotech.
6. 3 *M* Sodium acetate, pH 5.2.
7. 75% Ethanol.
8. 95% Ethanol.

2.2. DNA Fragmentation Using Hydrodynamic Shearing

1. HydroShear apparatus: GeneMachines, San Carlos, CA.
2. Disposable 1-mL syringe fitted with a 23-gauge needle.
3. 3 *M* Sodium acetate, pH 5.2.
4. 95% Ethanol.
5. 70% Ethanol.
6. 10X Polymerase buffer: 500 m*M* NaCl, 100 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 10 m*M* DTT.

2.3. DNA Fragmentation Using DNase I

1. DNase buffer: 50 m*M* Tris-HCl, pH 7.5, 1 m*M* $MnCl_2$.
2. DNase I: Sigma-Aldrich Co Ltd.
3. 500 m*M* Na_2EDTA, pH 8.0.
4. Phenol equilibrated to pH 8.0 with Tris buffer: Amersham Pharmacia Biotech.
5. 3 *M* Sodium acetate, pH 5.2.
6. 95% Ethanol.
7. 70% Ethanol.
8. 10X Polymerase buffer: 500 m*M* NaCl, 100 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 10 m*M* DTT.

2.4. DNA Shearing Using Sonication

1. Cup horn sonicator: Model W-385 Heat Systems Ultrasonics, Plainview, NY.
2. 10X Polymerase buffer: 500 m*M* NaCl, 100 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 10 m*M* DTT.

2.5. Fragment End Repair

1. 10X Polymerase buffer: 500 m*M* NaCl, 100 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 10 m*M* DTT.
2. T4 DNA polymerase: New England Biolabs.
3. 0.5 m*M* dNTPs: Amersham Pharmacia Biotech.

2.6. Fragment Size Selection

1. Low gelling temperature agarose: Life Technologies.
2. Submarine minigel apparatus, preferably 10 cm × 10 cm or smaller: Cambridge Electrophoresis, Cambridge, England.
3. 10X TBE: 108 g of Tris-base, 55 g of boric acid, 9.3 g of Na_2EDTA dissolved and made up to 1 L.
4. 5X Agarose gel sample loading buffer: 20% Ficoll 400, Amersham Pharmacia Biotech, 0.5X TBE.
5. 5X Agarose gel loading dyes: 20% Ficoll 400, Amersham Pharmacia Biotech; 0.5X TBE; 0.1% bromophenol blue; 1:2000 SYBR Green 1, FMC.
6. DNA size markers of a suitable range.
7. TE: 10 m*M* Tris-HCl, pH 8.0, 1 m*M* Na_2EDTA.
8. ß-Agarase 1: New England Biolabs.
9. Phenol equilibrated to pH 8.0 with Tris buffer: USB Amersham Pharmacia Biotech.
10. 3 *M* Sodium acetate, pH 5.2.
11. 95% Ethanol.
12. 70% Ethanol.
13. $TE_{0.1}$: 10 m*M* Tris-HCl, pH 8.0, 0.1 m*M* Na_2EDTA.

3. Methods

3.1. DNA Self-Ligation

1. Dissolve 3–5 µg of the linear DNA to be subcloned in 24 µL of $TE_{0.1}$ (*see* **Note 1**).

2. Add 3 μL of 10X ligase buffer and 3 μL of 10 m*M* ATP.
3. Add >2 U of T4 DNA ligase and incubate at 16°C for between 1 and 18 h. If the fragment has blunt ends, increase the enzyme quantity to approx 100 U and incubate the full 18 h.
4. Extract the DNA once with an equal volume of TE saturated phenol and remove the aqueous phase to a clean tube.
5. Add 0.1 vol of 3 *M* sodium acetate and 2.5 vol of 95% ethanol and chill on ice or at –20°C for 15 min.
6. Centrifuge for 15 min at 13,000*g* in a benchtop centrifuge at room temperature and carefully pour off the supernatant, trying to avoid dislodging the pellet. Wash the pellet with 1 mL of 70% ethanol and air-dry.

3.2. DNA Fragmentation Using Hydrodynamic Shearing

1. Carefully calibrate the system using a low total volume DNA solution (100–500 μL). When using a syringe and needle, try to exert a standard force while passing the DNA solution through the needle 1–6× and check the DNA sizes obtained (*see* **Note 2**). When using the HydroShear, follow the manufacturer's recommendation. Once the calibration is checked, the HydroShear will be extremely reproducible (*see* **Note 3**).
2. Dilute the sample DNA to the volume used during calibration, using polymerase buffer.
3. Shear the DNA exactly as during calibration.
4. If necessary, ethanol precipitate the DNA as described in **Subheading 3.1., steps 5** and **6**.
5. Redissolve the pellet in 30 μL of polymerase buffer.

3.3. DNA Fragmentation Using DNase I

Because fragmentation with DNase I is an enzymatic process, the degree of nicking is sensitive to DNA concentration, enzyme concentration and activity, temperature, buffer conditions, and duration of incubation. Each experiment should be preceded by calibration using conditions identical to those of the actual experiment (*see* **Note 4**). The conditions described here should be considered as a guide.

1. Dissolve 3–5 μg of circular or self-ligated DNA in 95 μL of DNase I buffer and place on ice.
2. Add 5 μL of stock DNase I to give a final concentration of 0.25 ng/100 μL.
3. Incubate the reaction at 15°C for around 10 min.
4. Add 1 μL of 500 m*M* Na_2EDTA to stop the reaction and immediately extract once with an equal volume of TE saturated phenol and remove the aqueous phase to a clean tube.
5. Ethanol precipitate the DNA as described in **Subheading 3.1., steps 5** and **6**.
6. Redissolve the pellet in 30 μL of polymerase buffer.

3.4. DNA Shearing Using Sonication

Sonicators break DNA by mechanical shearing. Large fragments shear much more readily than shorter ones, and the effect of this is to produce an initial rapid reduction of fragment size followed by progressive resistance to further breakage as the average fragment size is reduced. Even after prolonged sonication times, DNA of approx 150–200 bp will not appear to be reduced significantly in size, when assayed on agarose gels. The method described here uses a high-power cup horn sonicator that can be used to shear DNA indirectly, without direct contact of the probe with the DNA solution. Once the system is calibrated, using standard conditions, the sonicator can be used reproducibly (*see* **Note 5**). The conditions described here are for use with the Heat Systems W-385 fitted with a cup horn.

1. Fill the sonicator cup horn to a depth of around 3 cm above the probe with ice-cold water.
2. Dissolve 3–5 μg of self-ligated or circular DNA in 30 μL of polymerase buffer in a 1.5-mL microcentrifuge tube (*see* **Note 6**).
3. Using a small clamp or a water bath tube float, position the microcentrifuge tube approx 1 mm from the surface of the probe at approximately the center of its radius.
4. With the controller set at full power and 100% duty cycle, sonicate the sample for two bursts of 5 s (*see* **Note 7**). Between each burst, flick the tube to mix the contents and drive the liquid to the bottom of the tube (*see* **Note 8**).

3.5. Fragment End Repair

Whether the DNA was fragmented by DNase I, sonication, a French press or simply passing through a fine-gauge needle, the molecule ends will not be compatible with blunt-end cloning. Prior to any subcloning procedure, the DNA needs to be end repaired to remove any protruding 5' or 3' overhangs. This can be achieved simply and quickly by using T4 DNA polymerase, which possesses high 3'–5' exonuclease and 5'–3' polymerase activities.

1. To 3–5 μg of fragmented DNA dissolved in polymerase buffer add 4 μL of 0.5 m*M* dNTP solution.
2. Add 10 U of T4 DNA polymerase and incubate at a temperature of 15–20°C for 15 min (*see* **Notes 9–11**).
3. Heat-terminate the reaction at 70°C for 10 min.

3.6. Fragment Size Selection

1. Prepare a 0.6% low gelling temperature agarose 1X TBE minigel (a 10 cm × 10 cm gel requires approx 50 mL of gel) with a slot size of 1 cm and chill to 4°C to ensure complete setting of the gel and ease of comb removal (*see* **Note 12**).
2. To 30–40 μL of fragmented DNA add 6 μL of gel loading buffer and load the entire sample into a single well of the minigel. Apply suitable size markers (in loading buffer plus SYBR Green 1) to a well not immediately adjacent to the sample (*see* **Notes 13** and **14**).
3. Run the gel at <5 V/cm until the bromophenol dye of the markers has migrated only approx 2–3 cm.
4. Using a flat safety blade or scalpel, remove a gel slice corresponding to the size range(s) of fragments desired and place it in a 1.5-mL microcentrifuge tube (*see* **Note 15**). Ideally, the slice volume should be approx 250 μL; add TE if necessary (*see* **Note 16**).
5. Melt the slice at 65°C (around 10 min), making sure all of it has melted, and then cool it to 40°C.
6. Add 5 U of ß-agarase 1 and incubate at 40°C for 1 h (*see* **Note 17**).
7. Chill the tube on ice and briefly centrifuge it to pellet any undigested agarose. Remove the supernatant to a fresh tube.
8. Phenol extract and ethanol precipitate the DNA as described in **Subheading 3.1., steps 4–6**.

9. Redissolve the DNA in 50 µL of $TE_{0.1}$. This is sufficient for approx 50 small-scale ligations (10 ng of blunt vector) and should yield a few hundred to a few thousand recombinants per ligation.

4. Notes

1. It is possible to omit the self-ligation step, particularly where the DNA fragment is large. However, the subfragments will not be randomly distributed and the fragment ends will tend to be greatly overrepresented in the recombinant library.
2. When using a syringe and needle to shear DNA, it is possible to increase reproducibility by fabricating a weight-driven apparatus to standardise conditions.
3. The manufacturers of the HydroShear suggest that it is possible to proceed to cloning the sheared DNA without either end repair or size selection, greatly simplifying the process.
4. Each and every fragmentation experiment using DNase I needs to be carefully calibrated, as the degree of digestion is DNA concentration dependent and the DNase activity changes upon storage.
5. Although only a single calibration is needed for a given sonicator (provided the probe is not disconnected), it is important that the same supplier of tube is used, as, tube wall thickness may influence transmission efficiencies. It is also good practice to try to position the tube identically for each experiment.
6. Ligase buffer can be used during sonication thus avoiding an additional ethanol precipitation step . This may, however, have an effect on the specificity of breakage-factors such as salt concentration have not been fully evaluated in this regard but are likely to be significant.
7. Breakage by sonication is extremely severe. Try to minimize the duration of sonication as far as is possible. This may help to avoid internal damage to the DNA structure.
8. Conventional direct contact sonicators can be used for shotgun cloning and will usually require much reduced sonication conditions. Great care needs to be taken to clean the probe between experiments to avoid cross-contamination, and if the probe needs to be dismantled for cleaning, this may alter the calibration.
9. T4 polymerase end repair is carried out at or below room temperature because the enzyme exonuclease activity is elevated at temperatures approaching 37°C.

10. Short extension times are used for the end repair because it is possible to exhaust the supply of deoxynucleotides.
11. Other enzymes such as mung bean nuclease can be used for end repair, but little difference in subsequent cloning efficiency is observed.
12. Size exclusion columns such as the Chroma Spin from Clontech can be used as a simple alternative to fractionation through agarose for size-selection. However, they do tend to allow significant quantities of smaller fragments to pass through and unfortunately smaller fragments tend to clone much more readily.
13. DNA size markers are stained separately to avoid staining the sheared DNA. It may be necessary to run a small aliquot of the sheared DNA, also stained, in a separate well of the gel to check the size range.
14. An alternative to **Note 13** is to run both the markers and the sheared DNA unstained. The gel is then bisected vertically to separate the bulk of the sample but leaving a small edge piece within the half of the gel containing the markers. If the marker-containing half is then stained and the gel reassembled, both the size range and size distribution can be assessed.
15. Avoid ultraviolet light exposure of the fractionated sheared DNA during slice removal by masking the gel.
16. It is important to minimize the volume of the gel slice, both to minimize the quantity of agarose present and to keep volumes low so that subsequent clean-up procedures can be carried out in a single tube.
17. Many other procedures can be used for removing DNA from agarose ***(9)***. They tend to give similar yields of recovery. β-Agarase 1 is used in preference simply because it seems to work with the same efficiency every time and it is very simple to use.

References

1. Messing, J. and Bankier, A. T. (1989) The use of single-stranded DNA phage in DNA sequencing, in *Nucleic Acids Sequencing: A Practical Approach* (Howe C. J. and Waed E. S., eds.), IRL Press, Oxford, pp. 1–36.
2. Hong, G. F. (1982) A systematic DNA sequencing strategy. *J. Mol. Biol.* **158,** 539–549.
3. Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28,** 351–359.

4. Bankier, A. T., Weston, K. M., and Barrell, B. G. (1987) Random cloning and sequencing by the M13/dideoxynucleotide chain termination method, in *Methods in Enzymology*, vol. 155 (Wu, R. ed.), Academic Press, London, pp. 51–93.
5. Smith, V., Craxton, M., Bankier, A. T., Brown, C. M., Rawlinson, W. D., Chee, M. S., and Barrell, B. G. (1993) Preparation and fluorescent sequencing of M13 clones: microtitre methods, in *Methods in Enzymology,* vol. 218 (Wu, R. ed.), Academic Press, London, pp. 155–173.
6. Oefner, P. J., Hunicke-Smith, S. P., Chiang, L., Dietrich, F., Mulligan, J., and Davis, R. W. (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res.* **24,** 3879–3886.
7. Anderson, S. (1981) Shotgun DNA sequencing using cloned DNAse 1-generated fragments. *Nucleic Acids Res.* **9,** 3015.
8. Deininger, P. L. (1983) Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. *Anal. Biochem.* **129,** 216.
9. Sambrook, J., Frisch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

8

Cycle Sequencing

Robert W. Blakesley

1. Introduction

Cycle sequencing is a variant of standard dideoxy, chain terminator DNA sequencing ***(1–3)***. The protocol has gained popularity owing to at least three features: simple execution, robust performance, and signal amplification. The main characteristic that defines the cycle sequencing protocol is that the sequencing reactions are incubated in a thermal cycler, with programs similar to those used in polymerase chain reaction (PCR). This method assures efficient, reproducible utilization of even difficult templates by repeated thermal denaturation of the DNA template during the sequencing reactions. In fact, through heat cycling the same template molecules are used repeatedly, resulting in accumulation of sufficient sequencing signal even when a very limited amount of template is used, for example, when sequencing DNA from single bacterial colonies or phage plaques ***(4,5)***. It should be made clear, however, that no new template molecules are created as they are in PCR; cycle sequencing products accumulate linearly, not geometrically, in this single-primer DNA synthesis reaction. Cycle sequencing can be performed using a variety of labeling and detection techniques. The following protocol utilizes radioactive end-labeled primer and X-ray film.

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

The protocol starts with 5'-end labeling of the primer using T4 polynucleotide kinase and either ^{32}P- or ^{33}P-labeled ATP. After heat inactivation of the kinase, template DNA, *Taq* DNA polymerase and sequencing buffer are added, and the mixture is then distributed to four termination mixtures, each containing the four deoxynucleoside triphosphates (dNTPs) plus one specific dideoxynucleoside triphosphate (ddNTP). The thermal program is then started, typically for 20–30 cycles, repeatedly denaturing the template, annealing the primer, then synthesizing a complementary oligonucleotide until extension is terminated by incorporation of a ddNTP. In this way, double-stranded DNA is sequenced directly, avoiding the sometimes unreliable, inefficient, and time-consuming methods to prepare template in single-stranded form ***(3)***. Sequencing at high temperature reduces false stops induced by template secondary structures and improves the success with DNAs of a higher GC content. Also, background noise from primer bound to secondary sites is decreased as a result of the higher stringency of high-temperature primer hybridization and extension.

Sequencing with an end-labeled primer eliminates false signals caused by priming from nucleic acid contaminating the template, sometimes seen when labeled nucleotide is incorporated internally. Another advantage of using an end-labeled sequencing protocol is that sequence can be read starting several bases after the 3' end of the primer. Commercial availability of [γ-^{33}P]ATP provides the added benefits of increased band sharpness and read length ***(6)***. An alternative labeling strategy is to incorporate ^{33}P-labeled dideoxynucleotides during the sequencing reaction, which eliminates false bands resulting from premature termination ***(7)***.

Cycle sequencing requires that the DNA polymerase remain active after repeated exposure to 95°C. *Taq* DNA polymerase used here is thermally stable and, in addition, catalyzes DNA synthesis at 70°C. For many years, *Taq* DNA polymerase, available from a number of commercial sources, was the primary enzyme used in PCR and in cycle sequencing. Other thermally stable DNA polymerases may also work in cycle sequencing; however, optimal reaction conditions will vary from those provided in this protocol. A mutant of

Taq DNA polymerase ***(8,9)*** improves sequence quality by reducing the variation in band intensities. This enzyme accepts ddNTPs more readily than *Taq* DNA polymerase, requiring a different balance of ddNTP to dNTP in the reaction. Another thermal stable enzyme has shown good results in DNA sequencing ***(10)***. The *Taq*-based enzymes generally do not incorporate [α-^{35}S]dNTP efficiently, thereby generating weak signal, and offering no advantage for sequencing with internal labeling as is seen in the 37°C, single extension reactions of modified T7 DNA polymerase ***(11)***.

2. Materials

2.1. Endlabeling of Primer

1. Labeling buffer: 300 m*M* Tris-HCl, pH 7.8, 50 m*M* $MgCl_2$, 1 *M* KCl. Autoclave prior to use. Store at –20°C.
2. Kinase dilution buffer: 50 m*M* Tris-HCl, pH 7.6, 25 m*M* KCl, 5 m*M* dithiothreitol (DTT), 0.1 m*M* ATP, 0.2 mg/mL of bovine serum albumin (BSA), 50% (v/v) glycerol. To prepare, add separate filter-sterile solutions of DTT, ATP, and BSA to an autoclaved mixture of the salts and glycerol. Store at –20°C.
3. Sequencing primer (*see* **Note 1**), for example, for pUC18 and pUC19: 5'-CCCAGTCACGACGTTGTAAAACG-3' (23 bases), and T4 polynucleotide kinase are available commercially.
4. Labeled adenosine triphosphate (*see* **Note 2**): either [γ-^{32}P]ATP (≥ 3000 Ci/mmol, 10 mCi/mL) or [γ-^{33}P]ATP (1600 Ci/mmol), is available commercially.

2.2. Sequencing Reactions

1. Sequencing buffer: 300 m*M* Tris-HCl, pH 9.0 (25°C), 50 m*M* $MgCl_2$, 300 m*M* KCl, 0.5% (w/v) W-1 (Sigma, St. Louis, MO). Autoclave prior to use. Store at –20°C.
2. Formamide dye solution: 95% (v/v) deionized formamide, 10 m*M* EDTA, pH 8.0, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol. Store at –20°C.
3. 10X TE buffer: 100 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA. Autoclave prior to use. Dilute some TE buffer to 1X with autoclaved distilled water. Store at 4°C.

4. dNTP (dATP, dCTP, TTP), 7-deaza-deoxyguanosine triphosphate (7-deaza-dGTP) and ddNTP (ddATP, ddCTP, ddGTP, ddTTP) stock solutions, *Taq* DNA polymerase, cloning vectors (e.g., pUC19 DNA), and light silicone oil (optional) are available commercially.
5. 10X dNTP mix: Make a combined solution of 1 m*M* each dATP, dCTP, 7-deaza-dGTP, and TTP from the respective stock solutions in 1X TE buffer. Store at –20°C.
6. ddNTP working solutions: If not available commercially at this concentration, then make from the solid separate 10 m*M* working solutions of ddATP, ddCTP, ddGTP, and ddTTP in 1X TE buffer. Store at –20°C.
7. Sequencing mixes: Make 200-μL mixtures from the working 10X dNTP and the four working ddNTP solutions, according to **Table 1** (volumes are in microliters). Store at –20°C.

2.3. Equipment

1. Thermal cycler: A temperature cycling incubator capable of executing two consecutive programs over the temperature range of 45–95°C, with an incubation chamber for 0.2- or 0.5-mL microcentrifuge tubes and, preferably, with a heated lid.
2. Water baths or heater blocks: Capable of temperatures of 37, 55, and 70°C. Alternatively, the thermal cycler can be used for the single-temperature, timed incubations.
3. Microcentrifuge tubes: Polypropylene, 0.2 or 0.5 mL, with locking caps; caps must not open during the high-temperature incubations. Autoclave prior to use. For high sample throughput, a thermal cycler-compatible with 96-well plates is useful.
4. Automatic pipets: Capable of dispensing 0.5–20 μL and 10–100 μL.
5. Electrophoresis equipment and supplies for DNA sequencing ***(12)***.
6. X-ray film, single-sided with a blue base is recommended for optimal clarity and read length, for example, Kodak BioMax MR.

3. Methods

3.1. Endlabeling of Primer

1. One unit of kinase is sufficient to label 1 pmol of primer. Depending on the commercial source of enzyme, excess kinase may cause reduced labeling efficiency. It is recommended to dilute the concen-

Table 1
Preparation of Sequencing Mixes

	A	C	G	T
10X dNTP mix	20	20	20	20
10 m*M* ddATP	40	—	—	—
10 m*M* ddCTP	—	20	—	—
10 m*M* ddGTP	—	—	4	—
10 m*M* ddTTP	—	—	—	40
10X TE buffer	14	16	18	14
Autoclaved distilled water	126	144	158	126

Table 2
Weight of 1 pmol of Primer[a]

Primer length, b	Weight, ng
18	5.9
20	6.6
22	7.3
24	7.9
26	8.6
28	9.2
30	9.9

[a]Weight of 1 pmol is estimated by multiplying the primer length in bases by 0.33 ng.

trated enzyme solution to 1 U/μL with kinase dilution buffer before use, and use only 1 U/pmol of primer. The diluted enzyme is stable for several months when stored at –20°C.

Approximately 1 pmol of sequencing primer per set of sequencing reactions is end labeled using T4 polynucleotide kinase and fresh [γ-^{32}P]ATP or [γ-^{33}P]ATP (*see* **Note 2**). This reaction can be directly scaled for sequencing multiple templates with the same primer. If necessary, dilute the sequencing primer with 1X TE buffer to a final concentration of 0.5 pmol/μL (*see* **Table 2**).

2. Combine the following in order in an autoclaved microcentrifuge tube: 2 μL of 0.5 pmol/μL sequencing primer, 1 μL of labeling buffer, 2 pmol of radioactively-labeled ATP, and 1 μL of 1 U/μL T4 poly-

nucleotide kinase; total volume is 5 μL. Cap the tube and gently mix the contents. Collect the reaction volume by brief centrifugation in a microcentrifuge.
3. Incubate the reaction for 10 min at 37°C.
4. Incubate the tube for an additional 5 min at 55°C to terminate the labeling reaction. Centrifuge the tube briefly to collect the contents. Place the tube in wet ice. Although there is an excess of labeled ATP remaining, no cleanup of the end-labeled primer is required for performing cycle sequencing reactions. [γ-^{32}P]- and [γ-^{33}P]-labeled primers are useable for 1 wk and 3–4 wk, respectively, when stored at –20°C.

3.2. Sequencing Reactions

1. For convenience and to provide uniformity between the four dideoxy sequencing reactions, a prereaction mixture is first prepared for each template. Dilute the template to a concentration of approx 2 fmol/μL with 1X TE buffer. Compared to other methods, less DNA is introduced into cycle sequencing reactions (*see* **Notes 3–6**), 15–50 fmol being typical. The amount used depends on the size (length) of the DNA (see **Table 3**).
2. Prepare a prereaction mix by adding the following in order directly to the tube containing 5 μL of end-labeled primer: 4.5 μL sequencing buffer, 7–26 μL of ~2 fmol/μL DNA, 0.5 μL of 2.5 U/μL *Taq* DNA polymerase, and autoclaved distilled water to a total volume of 36 μL.
3. Cap the tube and gently mix the contents. Collect contents of the tube by brief centrifugation in a microcentrifuge. Place the tube in wet ice.
4. Label the caps of four microcentrifuge tubes A, C, G, and T; markings on tube walls usually come off during incubation. Prepare 10 μL reactions according to **Table 4** by placing the components in the appropriately labeled tubes (volumes in microliters).
5. Mix the contents of each tube gently. If using a thermal cycler without a heated lid, then oil must be added to completely cover the aqueous reactions; otherwise blank lanes in the gel may result. Dispense 20 μL of silicone oil on the surface of each reaction. Securely lock the caps on all four tubes. In either case, centrifuge briefly to collect volumes in the reaction tubes. Place the tubes in wet ice.
6. Start the thermal cycler on a program consisting of a soak for 3 min at 95°C to denature the template, followed by 20 cycles of denatur-

Table 3
Weight of Double-Stranded DNA Needed for One Reaction Set[a]

DNA length, kb	Weight, μg
0.5	0.016
1.0	0.033
3.0	0.10
5.0	0.12
7.5	0.14
10	0.17
25	0.31
45	0.50

[a]Weight of dsDNA ≤3 kb is estimated by multiplying the total DNA length in kb by 0.033 mg.

Table 4
Sequencing Reactions

Component	Tube A	Tube C	Tube G	Tube T
Prereaction mix	8	8	8	8
Sequencing mix-A	2	—	—	—
Sequencing mix-C	—	2	—	—
Sequencing mix-G	—	—	2	—
Sequencing mix-T	—	—	—	2

ation for 30 s at 95°C, annealing for 30 s at 55°C, and synthesis for 60 s at 70°C, and then followed by 10 cycles of denaturation for 30 s at 95°C and synthesis for 60 s at 70°C (*see* **Note 7**). When the temperature for the soak reaches 95°C, place the securely capped sequencing reaction tubes in the incubation chamber. If silicone oil is used in the tubes, blot the top of each tube with a wet paper towel to eliminate static electricity that may disrupt the two phases.

7. After completion of the program, about 2–2.5 h, remove the reaction tubes carefully from the incubation chamber. Terminate the reactions by adding 5 μL of formamide dye solution to each tube. Cap

the tubes, mix well, and collect the volumes by brief centrifugation. Store at –20°C. Analyze the reaction products by gel electrophoresis within 24 h for optimal results.

3.3. Gel Electrophoresis

1. Cycle sequencing reaction products are separated on a standard sequencing gel ***(12)***. To visualize from a few bases to several hundred bases from the 3' end of the primer, gels are prepared as 6% or 8% polyacrylamide in Tris–borate–EDTA buffer with urea as a denaturant. The bromophenol blue dye in the samples is allowed to reach within a few centimeters of the gel bottom before electrophoresis is terminated.
2. Heat denature the completed sequencing reactions at 70°C for 2 min. Collect the liquid by brief centrifugation of the tubes in a microcentrifuge.
3. Load 1–4 μL from each reaction in separate lanes of the sequencing gel. Store unused portions of the reactions at –20°C.
4. Following completion of electrophoresis, transfer the gel to chromatography paper. Dry the gel under vacuum prior to film exposure to increase band sharpness.
5. Expose single-sided X-ray film to the gel overnight at room temperature; an intensifying screen is not used.

4. Notes

1. Primer length, GC content, and the annealing temperature must be chosen carefully, as for PCR reactions ***(13)***. It is preferred to maintain the highest primer hybridization stringency so a minimum length of 20 bases and a GC content between 50% and 60% is recommended. Although not recommended, cycle sequencing with short primers (<20 mers) sometimes can be completed successfully by reducing the annealing temperature in a cycle from 55 to 45°C. Typically, reducing the lowest temperature of a cycle increases the chance of false priming and spurious sequence data. For sequencing AT-rich DNAs, utilize a primer that anneals to a region of locally high percent GC. A primer that binds to the 3' end of a homopolymer region avoids structural artifacts and eliminates out-of-register sequences ***(14)***. A primer with a few mismatches or a one-base dele-

tion relative to the template in the middle or 5' end can generate sequence, whereas a primer with a 3' end mismatch cannot.

2. [γ-^{33}P]ATP readily substitutes for [γ-^{32}P]ATP in the primer end labeling reaction, providing the advantages of sharper sequence images, longer shelf-life, and lower exposure hazard ***(6)***. Add 2 pmol of ^{33}P-labeled ATP to the end labeling reaction; no other changes are required. Typically, after a 16–24 h exposure the ^{33}P-labeled sequencing bands are easily visualized. [γ-^{35}S]ATP, on the other hand, is not recommended, as it is a very inefficient end labeling substrate. Such a reaction requires significantly more units of polynucleotide kinase and considerably longer incubation time, followed by extensive film exposure.
3. Sequencing data quality is directly related to template DNA purity. Even though cycle sequencing with end-labeled primer is more tolerant of contaminants, some background and false band artifacts may remain that are a result of template impurities. To maximize sequence yield, purify the DNA in large enough quantity or use commercial purification products in which other nucleic acids, proteins, and lipids are efficiently removed. Avoid introducing polymerase-inhibiting substances such as EDTA, sodium dodecyl sulfate (SDS) or salt to the reaction; clean up problem DNA using ammonium acetate/ethanol precipitation ***(15)***. Do not use Triton-containing solutions in DNA purification, as the carryover of Triton into sequencing reactions will contaminate the glass plates during electrophoresis, causing dark smears on the X-ray film. ATP-dependent DNase (Epicentre Technologies) treatment of a troublesome plasmid preparation is useful for removing spurious ssDNA and linear dsDNA from plasmids for cleaner sequencing results.
4. A great advantage of cycle sequencing is the signal increase, up to 10-fold, as result of the inherent linear amplification. This reduces the demand for scaleup of DNA preparations, especially of larger DNAs such as cosmids ***(16)***. Band intensity will increase with added DNA until the reaction becomes saturated at approx 500 fmol. Introducing more DNA rapidly depletes the nucleotide pool, creating a large amount of randomly terminated short oligos and indeterminable sequence. Similarly, a region of unusual nucleotide bias can unbalance the reaction, creating similar artifacts. To alleviate these problems, dilute a fresh aliquot of template 5–10-fold in 1X TE buffer and sequence again. Before starting any sequencing project, it is strongly recommended to perform agarose gel electrophoretic analysis to assess the quality and validate the quantity of a DNA preparation.

5. DNAs produced by PCR are readily sequenced by this method, even from one of the amplification primers ***(17–19)***. However, sequencing data quality is dependent on the specificity of the PCR; reactions containing multiple products or extensive truncated, failure products cannot be used. Multiple products must be separated from one another prior to sequencing, for example, by isolation through agarose gel electrophoresis. Sometimes, an internal or "nested" primer averts this problem ***(20)***. Alternatively, the specificity of the PCR can be optimized by redesign of the amplification primers or by using more stringent amplification conditions. An effective method to remove truncated or failure ssDNA products and excess primers is treatment of a portion of the reaction with 10 U of exonuclease I (Amersham) (Young and Blakesley, unpublished observations). For PCR products without these problems, purify an aliquot of the reaction from residual nucleotides, primers, and salts by selective precipitation using either isopropanol ***(21)*** or polyethylene glycol ***(22)***, or by glass matrix chromatography ***(17)***.
6. The attributes of cycle sequencing can be used advantageously to sequence DNAs isolated directly from single plasmid-containing bacterial colonies and bacteriophage plaques from an agar plate ***(4,5)***. The lysis supernatant (9 μL) is diluted to 26 μL with 1X TE buffer and cycle sequenced as described previously.
7. Operation of the thermal cycler affects the results of cycle sequencing similar to but to a lesser extent than those of PCR. Changes in the recommended program may be attempted to optimize results for a particular system, such as shorter run times, fewer cycles, or different annealing temperatures. For example, it was found that a minipreparation of plasmid DNA and the 23-base primer gave excellent results with a simplified program: 20 cycles of 30 s at 95°C, 60 s at 65°C with a temperature ramp of 30°C/min (Young and Blakesley, unpublished observations). In this example, however, a transition between temperatures faster than 30°C/min reduced the signal intensity and quality. Not all thermal cyclers perform equivalently under the same program. If the program fails to generate good sequence data, then use a program proven to generate PCR product.

References

1. Murray, V. (1989) Improved double-stranded DNA sequencing using linear polymerase chain reaction. *Nucleic Acids Res.* **17**, 8889.

2. Craxton, M. (1991) Linear amplification sequencing, a powerful method for sequencing DNA. *Methods Comp. Methods Enzymol.* **3,** 20–26.
3. Blakesley. R. W. (1991) Cycle sequencing, in *Methods in Molecular Biology, vol. 23: DNA Sequencing Protocols* (Griffin, H. G. and Griffin, A. M., eds.), Humana Press, Totowa, NJ, pp. 209–217.
4. Krishnan, B. R., Blakesley, R. W., and Berg, D. E. (1991) Linear amplification DNA sequencing directly from single phage plaques and bacterial colonies. *Nucleic Acids Res.* **19,** 1153.
5. Young, A. and Blakesley, R. (1991) Sequencing plasmids from single colonies with the dsDNA Cycle Sequencing System. *Focus* **13,** 137.
6. Pless, R. C. (1992) Cycle sequencing of DNA using 5'-^{33}P-labeled primers. *Focus* **14,** 102–103.
7. Fan, J., Ranu, R. S., Smith, C., Ruan, C., and Fuller, C. W. (1996) DNA sequencing with [α-^{33}P]-labeled ddNTP terminators: A new approach to DNA sequencing with Thermo Sequenase™ DNA polymerase. *BioTechniques* **21,** 1132–1137.
8. Tabor, S. and Richardson, C. C. (1995) A single residue in DNA polymerases of the *Escherishia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **92,** 6339–6343.
9. Vander Horn, P. B., Davis, M. C., Cunniff, J. J., Ruan, C., McArdle, B. F., Samols, S. B., et al. (1997) Thermo Sequenase DNA polymerase and T. acidophilum pyrophosphatase: new thermostable enzymes for DNA sequencing. *BioTechniques* **22,** 758–762.
10. Meis, R. (1997) Increased sensitivity in DNA sequencing with the new SequiTherm EXCEL™ II DNA polymerase. *Epicentre Forum* **4(2),** 5.
11. Tabor, S. and Richardson, C. C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
12. Bankier, A. T. (1991) Electrophoresis of sequence reaction samples, in *Methods in Molecular Biology, vol. 23: DNA Sequencing Protocols* (Griffin, H. G. and Griffin, A. M., eds.), Humana Press, Totowa, NJ, pp. 121–129.
13. Innis, M. and Gelfand, D. (1990) Optimization of PCRs, in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., eds.), Academic Press, San Diego, CA, pp. 3–12.

14. Khan, A. S., Wilcox, A. S., Hopkins, J. A., and Sikela, J. M. (1991) Efficient double stranded sequencing of cDNA clones containing long poly(A) tails using anchored poly(dT) primers. *Nucleic Acids Res.* **19,** 1715.
15. Crouse, J. and Amorese, D. (1987) Ethanol precipitation: Ammonium acetate as an alternative to sodium acetate. *Focus* **9(2),** 3–5.
16. Craxton, M. (1993) Cosmid sequencing, in *Methods in Molecular Biology, vol. 23: DNA Sequencing Protocols* (Griffin, H. G. and Griffin, A. M., eds.), Humana Press, Totowa, NJ, pp. 149–167.
17. Krishnan, B. R., Kersulyte, D., Brikun, I., Berg, C. M., and Berg, D. E. (1991) Direct and crossover PCR amplification to facilitate Tn*5supF*-based sequencing of λ phage clones. *Nucleic Acids Res.* **19,** 6177–6182.
18. Adams, S. M. and Blakesley, R. W. (1992) Sequencing a PCR-amplified DNA with the dsDNA Cycle Sequencing System. *Focus* **14**, 31.
19. Gyllensten, U. B., Allen, M., and Josefsson, A. (1992) Sequencing of in vitro amplified DNA, in *The PCR Technique: DNA Sequencing* (Ellingboe, J., ed.), Eaton, Natick, MA, pp. 1–15.
20. Phear, G. A. and Harwood, J. (1994) Direct sequencing of PCR products, in *Methods in Molecular Biology, vol. 31: Protocols for Gene Analysis* (Harwood, A. J., ed.), Humana Press, Totowa, NJ, pp. 247–256.
21. Brow, M. A. D. (1990) Sequencing with *Taq* DNA polymerase, in *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., eds.), Academic Press, San Diego, CA, pp. 189–196.
22. Paithankar, K. R. and Prasad, K. S. N. (1991) Precipitation of DNA by polyethylene glycol and ethanol. *Nucleic Acids Res.* **19,** 1346.

9

Cycle Sequencing of Polymerase Chain Reaction-Amplified Genomic DNA with Dye-Labeled Universal Primers

Johannes Zschocke and Georg F. Hoffmann

1. Introduction

Cycle sequencing of polymerase chain reaction (PCR) products has become the method of choice for sequencing known genomic DNA fragments. When fluorescent DNA sequencers are used, this requires the incorporation of a fluorescent dye into the cycle sequencing product either by using labeled sequencing primers or labeled terminators. The dye terminator approach is convenient because the same, unlabeled primers can be used for PCR amplification and cycle sequencing, and is cost efficient, particularly when a PCR product has to be sequenced a few times only. However, this method is not available for all DNA sequencers. Synthesizing specific dye-labeled primers, on the other hand, is expensive, and it is generally not affordable to label all PCR primers that at some stage one might like to use for sequencing.

Alternatively, a dye-labeled universal primer can be used for cycle sequencing if the appropriate recognition sequence is incorporated into the PCR product during the amplification reaction. This is easily achieved with PCR primers that contain an additional 17–20-mer tail

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

corresponding to the universal primer ***(1)***. Two different universal primers and accordingly two different PCR primer tails are necessary for sequencing in forward and reverse directions. The method has become widely used over the last few years, helped by significantly reduced primer costs and increased reliability of synthesizing 40-mer oligonucleotides.

Many laboratories routinely use universal M13 and M13 reverse tails for their forward and reverse PCR primers, respectively. We have opted for a choice of three cycle sequencing primers including M13 (but not M13 reverse), the sequences of which have been derived from the Bluescript SK vector (**Table 1**). The choice of primers was influenced by their estimated free energy of annealing and their melting temperatures ***(2)***. Having more than two universal sequencing primers enables us to choose tails for forward and reverse primers that contain least self- or cross-homologies, and this is always checked with one of the common computer programs (such as Oligo, MedProbe) when we design the 40-mers. We usually add three further bases (TAT) at the 5' end of the tail so that there is a degree of internal priming during cycle sequencing.

Cycle sequencing PCR products with labeled universal primers is cost efficient even if all PCR primers are routinely synthesized with a sequencing tail. The sequencing tail does not usually interfere with PCR amplification of genomic DNA; in contrary, in our experience the longmers sometimes work better than shorter PCR primers. Apart from labeled sequencing primers, standard reagents can be used for all steps. Sequencing several different samples at one time may be more convenient because only two or three sequencing primers are used and the cycling conditions are identical for all reactions. One major advantage over other sequencing methods is the earlier start of a good sequence. Regularly, the first 10–20 bases after the sequencing primer are not reliably read, and this can be a problem particularly when an exon of a gene has to be sequenced but only limited flanking intron information is available. In the present approach, the first 20 bases correlate to the specific part of the PCR primer, and the readable sequence usually starts well before the first unknown base.

Table 1
Dye-Labeled Universal Primers for the Cycle Sequencing Reaction[a]

M13	5' ¤-GTAAAACGACGGCCAGT
P51	5' ¤-AAGGGAACAAAAGCTGG
P172	5' ¤-TATAGGGCGAATTGGGT

[a] ¤, Position of the fluorescent dye label. The DNA sequence is incorporated as a 5'-tail into the initial PCR-primers.

2. Materials

1. PCR primers with 5' tails correlating to different universal primers, synthesized as described in **Subheading 3.1.**
2. Standard PCR reagents (e.g., platinum *Taq* DNA Polymerase with 10X PCR buffer, Life Technologies).
3. Standard agarose (e.g., SeaKem, FMC), 10X TBE buffer (e.g., BioWhittaker).
4. Purification columns for PCR products (e.g., MicroSpin S-300 HR columns, Amersham Pharmacia).
5. Universal cycle sequencing primers, 5'-labeled with fluorescent dye (Cy5 for use in ALFexpress DNA analysis system, Pharmacia).
6. Standard cycle sequencing reagents (e.g., Thermo-Sequenase kit, Amersham).

3. Methods

3.1. Design of Primers for Amplification of Genomic DNA (see Note 1)

1. Using standard criteria, choose a genomic DNA-specific primer pair for PCR amplification.
2. By checking for self- and cross-homologies, choose the best universal primer sequence to attach 5' to the genomic DNA-specific sequence. It is essential to use different universal primer sequences for forward and reverse primers.
3. Attach TAT in the 5' position to the universal primer sequence (optional). The resulting oligonucleotide should look as follows: 5'–TAT–universal primer sequence–genomic DNA-specific sequence–3'.

3.2. PCR Amplification of Genomic DNA and Cycle-Sequencing (see Note 2)

1. Perform standard PCR amplification of genomic DNA using primers designed as described previously. The annealing temperature in the first PCR cycles is usually 55 °C but may be raised to 65–72 °C after 5–10 cycles.
2. Check PCR product on a standard 2% agarose gel in 1X TBE buffer.
3. Purify PCR product using standard purification column, following a procedure recommended by the manufacturer. For the MicroSpin column (Pharmacia): Resuspend the resin in the column by vortex-mixing, loosen the cap, snap off the bottom closure, prespin the column for 1 min at 3000 rpm over a 1.5-mL reaction tube, place the column in a new 1.5-mL tube, remove and discard the cap, apply PCR product sample to the top center of the resin, and centrifuge for 2 min at 3000 rpm. The purified sample is collected in the bottom of the 1.5-mL support tube.
4. Perform standard four-tube cycle sequencing independently for forward and reverse directions using the appropriate universal primer that has been 5'-labeled with the appropriate fluorescent dye (**Table 1**).

4. Notes

1. The use of three different cycle sequencing primers allows six different possible combinations, at least one of which is usually satisfactory in terms of moderate or low degree of self- or cross-homologies. Occasionally, no acceptable universal sequencing primer combination can be found, and it may be necessary to design a new pair of genomic DNA-specific primers. It is useful to check for false priming sites of the universal primers within the sequence of the amplified PCR product, as this may interfere with cycle sequencing.
2. As always with cycle sequencing, the quality of the sequencing result depends to a great degree on the quality of the initial PCR product, which should be a single strong band on an agarose gel. The trouble-shooting recommendations are identical with those for standard PCR and cycle sequencing protocols.

References

1. Chadwick, R. B., Conrad, M. P., McGinnis, M. D., Johnston-Dow, L., Spurgeon, S. L., and Kronick, M. N. (1996) Heterozygote and

mutation detection by direct automated fluorescent DNA sequencing using a mutant *Taq* DNA polymerase. *BioTechniques* **20**, 676–683.

2. DeBellis, G., Manoni, M., Pergolizzi, R., Redolfi, M. E., and Luzzana, M. (1992) A more stringent choice of primers can improve the performance of fluorescent automated DNA sequencers. *BioTechniques* **13,** 892–897.

10

Automated Fluorescent DNA Sequencing on the ABI PRISM 377

Jonathan R. E. MacBeath, Suzanna S. Harvey, and Nicola J. Oldroyd

1. Introduction

Fluorescent automated DNA sequencing based on "one-lane, four-dye" technology has played a pivotal role in the success of the human genome project. One Applied Biosystems 377 sequencer can produce at least two runs per day of sequence reads averaging 750 bases in length, each run produces up to 96 reads. Improvements in sequencing chemistry (better polymerases and higher sensitivity dyes) have resulted in higher quality, more accurate sequencing data ***(1,2)***. This technology will be of enormous value in the next stage of genome sequencing, when the nature of sequence polymorphism in individual genes will be used to understand the complex relationships between genotype and phenotype. Fluorescent DNA sequencing is also rapidly becoming established as the technology of choice in basic research and clinical disciplines; methods for the molecular diagnosis of genetic disorders, pathogen identification, and genetic identification from biological specimens have all been developed and are being utilized in the field of DNA diagnostics ***(3–5)***.

In this chapter, we describe DNA sequencing using the ABI PRISM 377 sequencer. Besides eliminating the use of radioisotopes,

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

this instrument automates real time detection of DNA sequences, providing computer readable data that can be directly analyzed or entered into other programs, which can facilitate mutation detection, sequence editing, and assembly. For sequence analysis on the ABI PRISM 377, sequencing reactions are performed with fluorescent labels. Four different dyes are used to identify the A, C, G, and T extension reactions, which can be run on a single lane on a sequencing gel. This provides high throughput and unmatched precision by eliminating lane to lane mobility differences that can decrease base calling accuracy.

The 377 collects data at speeds of up to 200 bases per hour. It accepts plates of variable size which determines how far the DNA migrates before detection. These features together with careful consideration of gel formulation provides a straightforward way of achieving optimal electrophoresis and DNA resolution.

One of the most important factors that can dictate the success or failure of a DNA sequencing reaction is the quality of the template DNA. It is also the source of greatest variability between reactions. Protocols are presented for preparing DNA suitable for use as sequencing template for analysis on the Model 373, 377, and 310. We also describe how to prepare the glass plates and pour the polyacrylamide gel. Clean plates and high-quality reagents are crucial when using fluorescent detection methods, as dust and fluorescent contamination can cause high background levels, making data interpretation difficult (*see* **Note 1**).

Enzyme-mediated sequencing reactions are performed in a thermal cycler. This has the advantage of increasing the signal from a small quantity of template by utilizing multiple rounds of denaturation, primer annealing, and synthesis. Protocols for labeling DNA fragments using either 5' dye-labeled primers (dye primers [DP]) or 3' dye-labeled dideoxynucleotide triphosphate (ddNTPs) terminators (dye terminators [DT]) are presented. The labeling method used depends on sequencing strategy (*see* **Table 1**) and personal preference.

DP chemistry provides the user with even signal strength across all four bases but requires four tubes to complete the reaction owing to the attachment of the dye label to the extension primer. DT chem-

Table 1
DNA Sequencing Reagent Chemistry Recommendations[a]

	BigDye Terminators	dRhodamine Terminators	BigDye Primers
Comparative sequencing (germline mutations 50:50)	++	+	++
Comparative sequencing (somatic mutations 30:70)	+	+	++
Comparative sequencing (somatic mutations 10:90)	NR	NR	++
Homopolymer A or T >25 bp	NR	++	++
G–C-rich (>65%) templates	++	+	++
A–T-rich (>65%) templates	++	++	++
BAC, Cosmid, λ Templates	++	+	+

[a]This table has three levels of recommendations: recommended (++), satisfactory (+), and not recommended (NR). The definitions of these recommendations are as follows: Recommended: This chemistry is the best choice most of the time. Satisfactory: This chemistry will work. You should be able to get good results, but there is a better choice of chemistry for the given application. Not recommended: This chemistry might work. In most cases, the results will be suboptimal in terms of read length and/or accuracy.

istry is the more popular choice for the majority of applications because of the single-tube reactions and less hands on time required.

All reagent kit formulations contain the sequencing enzyme Ampli*Taq* DNA Polymerase FS (Ampli*Taq* FS) and is considered by many to be the ideal enzyme for fluorescent sequencing. Ampli*Taq* FS is a variant of *Thermus aquaticus* DNA polymerase, which contains a point mutation at the active site. This results in reduced discrimination against fluorescent ddNTPs which leads to a more even peak intensity and easier removal of unincorporated ddNTPs (DT chemistry only). The enzyme has a second mutation in the amino terminal domain that virtually eliminates the 5' to 3' exonuclease activity of Ampli*Taq* DNA polymerase. Sequencing kits have been formulated with a thermally stable pyrophosphatase to eliminate problems associated with pyrophosphorolysis (DP chemistry only).

The use of dichlororhodamine (dRhodamine) terminators provide more even peak heights than the original rhodamine dye termina-

tors, these dyes have narrower emission spectra giving less spectral overlap and therefore less noise. Sequence data also show a reduction of the weak G after A pattern characteristic of the original rhodamine terminators *(1,2)*. The improved peak eveness allows greater accuracy in base calling, longer reads, and the ability to use dye-labeled terminators for heterozygote analysis *(6)*.

BigDye Terminators offer the same benefits as dRhodamine terminators, and in addition employ higher sensitivity dyes. This represents the latest development in DT chemistry, and is the chemistry of choice for many users. The new BigDye structures contain a fluorescein donor dye linked to a dRhodamine acceptor dye. The excitation maximum of each dye label is that of the fluorescein donor and the emission spectrum is that of the dRhodamine acceptor. The linker allows extremely efficient energy transfer between the donor and acceptor dyes; hence, the BigDye Terminators are two to three times brighter than the standard dye terminators when incorporated in cycle sequencing products. They also have narrower emission spectra, giving less spectral overlap and therefore less noise. The brighter signal and decreased noise provide an overall four to fivefold improvement in sensitivity.

BigDye Primers utilize the same type of high-sensitivity dye combination as applied to the BigDye Terminator chemistry. These kits have been formulated with 7-deaza GTP to minimize band compressions, and nucleotide mixes have been optimized to give a good signal balance up to 1000 bases. As a result of improved signal strength, BigDye Primers/Terminators enable direct sequencing of genomic and BAC DNA templates *(2,7)*, and reduced volume reactions can also be done on some templates including polymerase chain reaction (PCR) products and plasmids (*see* **Note 2**).

Other methods are also emerging for sequencing large or difficult to sequence DNA fragments. Applied Biosystems have made available a transposon-based sequencing method. Transposon insertion for DNA sequencing and mapping provides a superior alternative to methods such as shotgun sequencing and primer walking *(8)*.

This chapter provides detailed instructions for performing sequencing on the ABI PRISM 377 running version 2.5 data collec-

tion (or higher) and version 3.3 sequence analysis (or higher), using dRhodamine Terminators, BigDye Terminators, and BigDye Primers. Further information can be found in the appropriate 377 user manual, DNA sequencing chemistry guide (Applied Biosystems; P/N 4305080), and protocol booklet supplied with the sequencing kit.

2. Materials

The operation of the ABI PRISM 377 is fully covered by the chemistry and instrument user manuals, as are the recommended sources for reagents and ancillary equipment needed for their successful use. A summary of the items needed, with important points to consider, is given here.

1. Water: For all reagents, gels, and buffers, use 18 MΩ resistance or better, minimal organic content.
2. Acrylamide: Bio-Rad or Amresco (19:1 acrylamide–bis-acrylamide) is recommended, as preweighed powder or as a premixed 40% solution. Stock solution should not be kept for more than 1 mo, at 4°C.
3. Tris-base, boric acid, Na_2EDTA, urea, and ammonium persulfate: Analar grade from major lab suppliers.
4. TEMED: Bio-Rad or Amresco.
5. Formamide: Bio-Rad or Amresco. Deionize with mixed bed resin (Bio-Rad AG501-X8) for 10–15 min and store frozen as single-use aliquots.
6. Centricon-100 microconcentrator columns: Amicon.

The following items can be ordered from Applied Biosystems:

7. BigDye Terminator ready reaction kit (100 reactions, P/N 4303149).
8. dRhodamine Terminator ready reaction kit (100 reactions, P/N 403044).
9. BigDye Primer ready reaction kit (100 reactions; -21 M13, P/N 403051; M13 reverse, P/N 403052).
10. 25 mM EDTA, pH 8.0, 50 mg/mL blue dextran (1 mL, P/N 402055).
11. ABI PRISM miniprep kit for plasmid purification (100 purifications, P/N 402790).
12. Ampli*Taq* Gold 5 U/µL plus 10X PCR buffer II–25 m*M* $MgCl_2$ solution (250 U, P/N N8080241).

Some items are essential and some are optional, depending on the DNA template (e.g., PCR product or plasmid) and sequencing chemistry used (dye primer vs dye terminator). All other reagents should be of analytical or molecular biology grade.

It is assumed that the user's laboratory will have all the general equipment necessary for molecular biology work. The sequencing protocols provided (*see* **Subheading 3.2.**) were optimized using Applied Biosystems thermal cyclers. Reactions can also be carried out on other manufacturer's cyclers, but the cycling parameters may need to be optimized. Ramping time is very important. If the thermal cycler is ramping too fast (>3°C/s), poor (noisy) data may result.

3. Methods

3.1. Preparation of Template DNA

3.1.1. Host Strains

The host strain from which template DNA is isolated can have a profound effect on sequence data quality. It has been reported that endonuclease A (*end*A$^+$) *Escherichia coli* strains yield nicked plasmid DNA that may produce smearing on manual sequencing gels ***(9)***. A similar, but not perfect, correlation has been found for templates prepared for automated fluorescent sequencing (A. Blasband, Applied Biosystems). Strains DH5α (*end*A$^-$) and HB101 (*end*A$^+$) have been found to consistently yield the best quality DNA for sequencing. Strains JM109 (*end*A$^-$), XL1-Blue (*end*A$^-$), and MV1190 (*end*A$^+$) frequently yield sequencable DNA while strain JM101 (*end*A$^+$) rarely does.

3.1.2. DNA Preparation Kits

DNA preparation methods for a number of templates including bacteriophage, plasmids, cosmids, and bacterial artificial chromosomes (BACs) are now available in kit form ***(10)***. These can be divided into two types based on the purification resin used.

Kits based on ion-exchange resins give excellent results providing they are not overloaded. DNA is eluted from the resin in a high

concentration salt solution. This should always be followed by a desalting step to prevent salt carryover into the cycle sequencing reaction (*see* **Note 3**).

Kits based on silica resins can give variable results. DNA is selectively bound in a high-concentration chaotropic salt solution and is eluted from the resin in water. The purified DNA may still contain salt so a desalting step should be added for consistency. Care should be taken not to allow silica finings to pass through into the DNA preparation, as silica inhibits the DNA polymerase in the subsequent cycle sequencing reaction (*see* **Note 3**). Given these problems, Applied Biosystems and others supply miniprep kits for the purification of plasmid DNA that have been validated for fluorescent sequencing.

3.1.3. Purification of Plasmid and Cosmid DNA

1. With a sterile toothpick or inoculating loop, collect a small sample of bacteria from a well-separated colony and rinse the cells into 5 mL of Terrific Broth in a 50-mL polypropylene screw-capped tube. Shake the culture overnight at 37°C.
   (To prepare Terrific Broth, add 1 mL of a sterile solution of 0.17 *M* KH_2PO_4, 0.72 *M* K_2HPO_4 to 9 mL of base broth (base broth: 12 g of Bacto-tryptone, 24 g of Bacto-yeast extract, 4.0 mL of glycerol, q.s. to 900 mL with deionized water [dH_2O] and then autoclave).
   The quantity of DNA isolated from JM101 can be improved by chloramphenicol amplification of the plasmid. For this procedure, the culture should be grown in 5 mL of LB (10 g of Bacto-tryptone, 5 g of Bacto-yeast extract, 10 g of NaCl, per liter). Shake the culture at 37°C until the absorbance at 600 nm is between 0.45 and 0.7. At that point, add 35 µL of 34 mg/mL chloramphenicol and continue incubation of the culture overnight at 37°C.
2. Pellet the entire culture in a single tube by successive 1-min centrifugations of 1.5-mL aliquots.
3. Remove the final supernatant by aspiration. By pipeting the solution up and down, resuspend the cell pellet in 200 µL of GTE buffer (50 m*M* glucose, 25 m*M* Tris-base, 10 m*M* EDTA, pH 8.0).
4. Add 300 µL of freshly prepared 0.2 *N* NaOH, 1% sodium dodecyl sulfate (SDS). Mix the suspension by inverting the tube several times and incubate on ice for 5 min.

5. Neutralize the solution by adding 300 μL of 3.0 *M* potassium acetate, pH 4.8, mix by inverting the tube several times and incubate on ice for 5 min.
6. Remove cellular debris by centrifuging for 10 min at room temperature, and then transfer the supernatant to a clean tube.
7. Add RNase A (DNase free) to a final concentration of 20 μg/mL and incubate the tube at 37°C for 20 min.
8. Extract the sample twice with 400 μL of chloroform. (Mix the solution by hand for 30 s for each extraction. Centrifuge the tube for 1 min to separate the phases and transfer the upper aqueous phase to a clean tube between extractions.)
9. Precipitate the total DNA by adding an equal volume of 100% isopropanol and immediately centrifuge the tube for 10 min at room temperature. Wash the DNA pellet with 500 μL of 70% ethanol and then dry under vacuum for 3 min.
10. Dissolve the pellet in 32 μL of dH_2O, and precipitate the plasmid DNA by first adding 8 μL of 4 *M* NaCl, and then 40 μL of autoclaved 13% polyethylene glycol $(PEG)_{8000.}$ Mix the sample thoroughly and incubate on ice for 20 min.
11. Pellet the DNA by centrifugation for 15 min at 4°C in a fixed-angle rotor. It is important to adhere to the recommended temperature of 4°C at this step.
12. Carefully remove the supernatant by aspiration and resuspend the pellet in 450 μL of sterile dH_2O.
13. Extract the sample twice with chloroform.
14. To the final aqueous layer placed in a clean tube, add 40 μL of 3 *M* sodium acetate, pH 5.5 and 1 mL of 95% ethanol. Mix thoroughly and set the sample on ice for 20 min.
15. Centrifuge the sample in a microcentrifuge for 25 min. Rinse the pellet with 500 μL of cold 70% ethanol.
16. Dry the pellet under vacuum for 5 min and resuspend the DNA in 20 μL of dH_2O. Store the purified DNA at –20°C.

This procedure will typically yield from 5 to 30 μg of DNA per 1.5 mL of culture.

3.1.4. Template Production Using PCR

For primer design considerations, *see* **Subheading 3.1.7.**

A typical 50-μL PCR reaction would contain 10–100 ng of DNA, 1.25 U of *Taq* DNA polymerase, 200 μ*M* dNTPs and 10 pmol each of the forward and reverse primer. A suitable PCR reaction buffer should be used. It may be necessary to optimize magnesium chloride concentration, and we strongly suggest that this is determined empirically for each primer set to achieve optimal PCR performance ***(11)***.

If necessary, overlay each reaction with 80 μL of light mineral oil. For initial denaturation, samples are placed in the preheated block and incubated at 95°C for 1 min. With an Applied Biosystems instrument, a typical thermal cycling protocol is as follows:

DNA Thermal Cycler TC1 or 480	9600, 9700, or 2400 Thermal cycler
94°C for 1 min	94°C for 15 s
55°C for 1 min	55°C for 15 s
72°C for 1 min	72°C for 15 s
for 20–30 cycles	for 20–30 cycles

The cycling parameters you choose will largely be determined by the annealing characteristics of your primers. With well-designed primers, an annealing temperature of 5°C below the T_m for your primers will work.

Use of Hotstart PCR can greatly facilitate direct sequencing of PCR products by improving the specificity and sensitivity of the reaction. By using Ampli*Taq* Gold in place of standard *Taq* DNA polymerase, effects such as mispriming and primer–dimer formation can be minimized, thus dramatically improving the quality of the sequencing data ***(11,12)***.

3.1.5. Purification of PCR Products

PCR products should be purified to remove excess reagents or secondary products produced during the reaction. Removal of the PCR primers prior to DNA sequencing is advisable because these could compete with the DNA sequencing primers. Removal of the unincorporated dNTPs from the reaction is also advisable, as their presence could alter the dNTP/ddNTP ratio of the DNA sequencing reaction. There are three methods to consider.

The first involves passing a PCR reaction mix through a Centricon-100 column. This removes any reactants <100,000 Dalton, and equates to a retention cutoff (bp) of >125 for double-stranded DNA. This method provides a useful means for primer–dimer removal ***(13)***. To purify PCR fragments with Centricon-100 columns:

1. Assemble the Centricon-100 column according to the manufacturer's recommendations.
2. Load 1 mL of dH_2O onto the column.
3. Add the entire PCR sample to the column.
4. Centrifuge the column at 3000*g* in a fixed-angle centrifuge for 5 min, to remove small molecular weight material (e.g., dNTPs, primers, and salt).
5. Add a further 1 mL of dH_2O to the column and recentrifuge for 5 min.
6. Remove the waste receptacle and attach the collection vial.
7. Invert the column and spin it at 270*g* for 2 min to collect your sample. This should yield approx 40–60 μL of sample.

The second method employs exonuclease I and shrimp alkaline phosphatase to treat PCR reactions. These enzymes degrade residual PCR primers and dephosphorylate the residual dNTPs. This method is easy to use, cost effective, and amenable to high sample throughput, but will not eliminate any secondary products (e.g., primer–dimer) from the PCR reaction ***(13)***.

Gel purification allows isolation of the fragment of interest from contaminating secondary products, and can be used with suboptimal PCR conditions. However, results can be variable and visualization of the gel with UV light can result in "nicking" of DNA. Only high-quality low melting point agarose should be used for gel purification ***(13)***.

3.1.6. Template Quantitation

Accurate quantitation of the DNA template for cycle sequencing is important for good, even signal strength. DNA can be quantitated by agarose gel electrophoresis followed by ethidium bromide staining and comparison of the fluorescence intensity under UV light to a control of known concentration. DNA quantitation by absorbance

is highly variable owing to sensitivity problems and levels of contamination from protein, RNA, chromosomal DNA, and salt.

3.1.7. Primer Design and Storage

Oligonucleotide primers for PCR should be designed so that they hybridize to opposite strands flanking the region of interest in the target DNA. They should be between 20 and 30 bases in length with melting temperatures (T_m) of 50–70°C. If possible, primers with similar T_m should be created. Primers should be made with a GC content of 50–55%, and long runs of a single base (that is, more than three or four, particularly G or C) should be avoided. The primers should not contain regions of complementarity either to themselves or to other primers in the reaction. Complementarity at the 3' ends should especially be avoided as this may lead to primer–dimer formation. Use of good primer design software is recommended.

Primer stock solutions should be stored in aliquots at –20 or –70°C in a noncycling freezer. Working solutions can be stored for 2 wk at 4°C.

Primer design considerations for sequencing reactions are very similar to those described previously. Primers with low (<45°C) and high (>65°C) melting temperatures do not always produce good results.

3.1.8. Thermal Cycler Considerations

For optimal results in both PCR and cycle sequencing, the thermal cycler must deliver reproducible thermal profiles at every well. Cycler performance can be checked by running PCR and cycle sequencing control reactions regularly and checking for changes in the overall length of the cycling time and quality of the results.

3.2. Sequencing Protocols

The following list gives recommended DNA concentrations and quantities for each sequencing chemistry. The ranges given in the list should work for all primers. The amount of template DNA, PCR product, or plasmid required will also depend on the size (bp) and purity of the product. In general, higher DNA quantities give higher signals.

DNA	dRhodamine terminators	BigDye terminators	BigDye primers
Single-stranded	50–100 ng	50–100 ng	300–600 ng
Plasmid DNA	200–500 ng	200–500 ng	600–1200 ng
PCR product	30–90 ng	30–90 ng	30–90 ng
BAC	N/A	300–600 ng	300–600 ng

The type of tubes required for the cycle-sequencing reactions will depend on the type of thermal cycler used. For the DNA Thermal Cycler (TC1) and the DNA Thermal Cycler 480, use 0.5-mL GeneAmp Thin-Walled PCR tubes. For the GeneAmp Systems 9600, 9700, and 2400 use 0.2-mL MicroAmp reaction tubes.

3.2.1. Cycle Sequencing with dRhodamine and BigDye Terminators

For standard templates (e.g., PCR products and plasmids), prepare the following reaction mixture. For each 20-μL reaction add the following reagents to a separate tube.

Terminator ready reaction mix	8 μL
Template	*See* above **(Subheading 3.2.)**
Primer	3.2 pmol
Deionized water	q.s. to 20 μL

Mix the reagents well and centrifuge briefly.

The following cycle sequencing profiles have been optimized for all Applied Biosystems thermal cyclers and consumables. If you use different thermal cyclers or reaction tubes, you may need to optimize thermal cycling conditions. If using the TC1 or 480 PCR instruments overlay the reaction mixture with 40 μL of light mineral oil. Place the reaction tubes in the thermal cycler preheated to 96°C and set the volume to 20 μL. Repeat the following for 25 cycles.

DNA Thermal Cycler TC1 or 480	9600, 9700, or 2400 Thermal cycler
96°C for 30 s	96°C for 10 s
50°C for 15 s	50°C for 5 s
60°C for 4 min	60°C for 4 min

Rapid thermal ramp to 4°C and hold until ready to purify. Rapid thermal ramps between temperatures (1°C/s) should be programmed on the cycler. Following cycle sequencing, centrifuge down the contents of the tubes and proceed to purifying extension products.

Cycle sequencing reactions can also be carried out on the CATALYST 800 Molecular Biology Labstation or the ABI PRISM 877 Integrated Thermal Cycler. Refer to the appropriate chemistry guide or instrument manual for further details.

With BigDye chemistries reduced volume reactions can be run on some templates (e.g., PCR products and plasmids). A 5X sequencing buffer is available from Applied Biosystems (P/N 4305605), for dilution of the ready reaction mix prior to cycle sequencing (*see* **Note 2**).

3.2.2. Purification of Extension Products

For this section, the purification methods described are appropriate for dRhodamine and BigDye Terminator chemistries. With the following procedures, traces of unincorporated terminators may be seen at the beginning of the sequence data, but this is usually minimal ***(14)***. For ethanol–sodium acetate precipitation repeat the following:

1. For each 20 μL sequencing reaction, prepare a 1.5-mL microcentrifuge tube containing:
   2.0 μL of 3 *M* sodium acetate, pH 4.6.
   50 μL of 95% ethanol at room temperature.
2. Pipet the entire contents of each extension reaction into a tube of sodium acetate–ethanol mixture. Mix thoroughly. If using reactions carried out on the TC1 or 480 thermal cyclers, remove the reaction carefully, taking care to avoid carryover of the oil.
3. Vortex-mix the tubes and leave at room temperature for 15–20 min to precipitate the extension products. Precipitation for longer may encourage carryover of excess dye molecules.
4. Centrifuge the tubes in a microcentrifuge for 20–30 min at maximum speed, noting the tube orientation.
5. Carefully aspirate the supernatant with a separate pipet for each sample and discard. At this point, a pellet may or may not be visible.

Respin the tubes and remove the final traces of ethanol with a fine bore pipet.

6. Rinse the pellet with 250 µL of 70% ethanol. Centrifuge for 5 min at maximum speed, again taking care to remove all of the ethanol.
7. Dry the pellet in a vacuum centrifuge for approx 5 min, or until dry. Do not overdry.

Simplified ethanol precipitation procedures and alternative purification procedures are available utilizing spin columns, refer to the appropriate protocol or chemistry guide for further details.

3.2.3. Cycle Sequencing with BigDye Primers

Two kit formulations incorporating -21 M13 or M13 reverse primers currently exist for sequencing with BigDye Primers. If required, these primer binding sites can be incorporated at the end of PCR products ***(13)***.

For standard templates (e.g., PCR products and plasmids), prepare the following reaction mixture. Aliquot the reagents into four separate PCR tubes. Refer to the list in **Subheading 3.2.** for recommended DNA concentrations.

Reagent	A (µL)	C (µL)	G (µL)	T (µL)
Ready reaction mix	4	4	4	4
DNA template	1	1	1	1
Total volume	5	5	5	5

Mix reagents well and centrifuge briefly, if using the DNA thermal cycler (TC1 or 480) add 20 µL of light mineral oil. Place the tubes in a thermal cycler preheated to 96°C and set the volume to 5 µL.

DNA Thermal Cycler TC1 or 480	9600, 9700, or 2400 Thermal cycler
95°C for 30 s	96°C for 10 s
55°C for 30 s	55°C for 5 s
70°C for 1 min	70°C for 1 min
Repeat for 15 cycles	Repeat for 15 cycles
95°C for 30 s	96°C for 10 s
70°C for 1 min	70°C for 1 min
Repeat for 15 cycles	Repeat for 15 cycles

Rapid thermal ramp to 4°C and hold until ready to purify.

Rapid thermal ramps between temperatures (1°C/s) should be programed on the cycler. Following cycle sequencing centrifuge the contents of the tubes. To concentrate the samples for loading:

1. Add 53 μL of 95% ethanol to a clean microcentrifuge tube. (**Note:** The use of sodium acetate is not necessary for precipitation.)
2. Pipet the extension reactions from the bottom of each of the four tubes into the ethanol and mix thoroughly. If using reactions carried out on the TC1 or 480 thermal cyclers, remove the reaction carefully taking care to avoid carryover of the oil (*see* **Note 4**).
3. Leave at room temperature for 10–15 min to precipitate the extension products.
4. Centrifuge the tube in a microcentrifuge for 10–20 min at maximum speed. Carefully aspirate the supernatant and discard. At this point, a pellet may or may not be visible.
5. As an optional step, the pellet can be rinsed with 250 μL of 70% ethanol and recentrifuged for 5 min. Again, carefully aspirate the supernatant and discard. Dry the pellet in a vacuum centrifuge for 1–3 min, or until dry. Do not overdry.

3.3. Preparing Acrylamide Denaturing Gels

For consistent results, reagents of the highest possible purity should be used (*see* **Note 5**). Be sure to wait at least 2 h after casting to ensure complete polymerization, but do not wait longer than 6 h.

3.3.1. Preparation of Stock Solutions

Some of the chemicals used are potentially hazardous; gloves should be worn at all times and work carried out in a fume hood when handling acrylamide and urea solutions (*see* **Note 6**).

1. 10X TBE (per liter): Tris-base (108.0 g), boric acid (55.0 g), $Na_2EDTA.2H_2O$ (8.3 g). Working solution (1X) is 89 m*M* Tris-base, 89 m*M* Boric acid, 2 m*M* EDTA; pH 8.3 at ambient temperature.
2. 1X TBE: Dilute 140 mL of 10X TBE with dH_2O to 1400 mL. TBE buffer should be discarded if the pH is not 8.3 (± 0.2). Do not attempt to adjust the pH.
3. 40% (19:1) acrylamide stock solution: 38.0 g of acrylamide, 2.0 g of *bis*-acrylamide. Dissolve in enough dH_2O to bring the total volume

up to 90 mL. Add 10 g of mixed-bed ion-exchange resin (Bio-Rad AG501-X8), and stir at room temperature for 10 min. Filter solution through a 0.2-μm cellulose nitrate filter, and adjust the volume to 100 mL with dH_2O. This solution is stable for approx 1 mo when stored at 4°C.

In addition to these reagents, ensure that you have each item on the following list ready before beginning to mix a sequencing gel: urea, *N,N,N′,N′*-tetramethylethylenediamine (TEMED), 10% (w/v) ammonium persulfate in dH_2O (prepare fresh daily, and store at 4°C) and mixed-bed ion-exchange resin.

3.3.2. Preparing the Plates and the Gel Cassette

This subheading describes pouring a gel with the gray cassette and gel pouring fixtures. Using this is recommended as it simplifies pouring and provides greater reproducibility ***(15)***.

1. To wash the glass plates: Prepare a 1% Alconox (Aldrich) solution in deionized (MilliQ) water. Fill a wash bottle with the solution and use a small amount to remove any grease or remaining gel particles from the plates. Lint-free tissue should be used to wipe the plates, be careful not to scratch or abrade the glass surfaces. Rinse the plates thoroughly (in a direction away from the read region) with dH_2O to remove all traces of alconox and air-dry. Each time you use the glass plates, use them with the same side of each plate on the inside.
2. To mount the glass plates in the cassette: Lay the gel cassette on a flat surface and lift the laser beam stop at the bottom of the cassette. The plates are etched on the outside bottom corner. Place the rear, noneared plate into the cassette with the inside of the plate facing up. Make sure the inside of the plate is free of water droplets, dust, lint, or anything else that may fluoresce or scatter light. Place the 0.2-mm gel spacers on the outer edges of the rear plate. The V-shaped cutout at the top of each spacer should point to the inside of the plate. Align the spacers at the bottom so that they are flush to the inside of the beveled edge of the glass plate. (Optionally, use a small amount of dH_2O to help anchor the spacers in place.) Place the notched (front) plate on top of the plain (rear) plate and spacers with the inside facing down. Align the plates so that the bottom ends are flush and with the notch on the front plate orientated toward the top of the cassette (*see* **Fig. 1**). Push the assembled plates toward the bottom of the

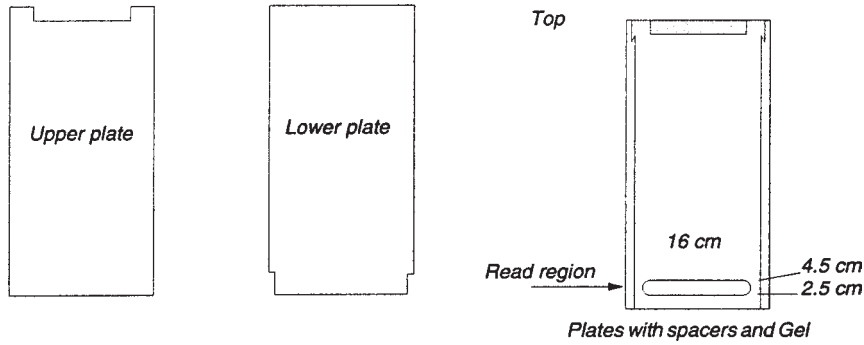


Fig. 1. Glass plates aligned for placement in the cassette.

cassette. Ensure that the cassette stops are firmly against the upper edge of the rear plate notches, then lock the plates into position with the cassette clamps. Check the alignment of the bottom ends of the plates to ensure that no slippage occurred during clamping. Realign, if necessary.

3. To attach the gel injection device: Open the top pair of cassette clamps. Place the top clamp on the top edge of the glass plates and slide the top pegs over the ears of the glass plates. The side pegs of the clamp fit into the clamp area on the cassette. Insert the casting comb and apply a small amount of pressure to the top clamp over each ear. Close the cassette clamps. Ensure that the top clamp is seated properly over the ears of the glass plates to prevent leaking. Verify the tightness of the comb by gently pulling on the teeth. If the comb is too loose, verify the top clamp is seated properly over the ears of the glass plates or use another comb. If the comb is still too loose, then replace the top clamp with binder clips during gel casting (*see* **Subheading 3.3.4.**). Even if the comb is too loose, the top clamp must remain clamped in position on the gel cassette to attach the gel injection device properly. Insert the orange gasket (ridge side up) in the gel injection device, if one is not present. (When inserting the gasket, make sure the luer fitting [hole] in the gasket is lined up with the hole in the injection device, and that no folds or irregularities in the gasket are present). To attach the injection device, open the bottom clamps on each side of the gel cassette and slide the injection device over the ends of the glass plates until the plastic tabs of the device snap into the cutouts in the gel cassette. Close the cassette clamps. The cassette is now ready for you to inject the gel.

3.3.3. Preparing Acrylamide–Urea Gel Solutions

The following recipe describes the volumes necessary for 4.25% acrylamide, 6 *M* urea, 1X TBE gel solution. For a 36-cm well-to-read (WTR) gel, prepare 50 mL of solution. For a 48-cm (WTR) gel, prepare 80 mL of solution. These amounts allow for some spillage. With experience, you may be able to prepare less solution.

Combine the following components in a 150-mL beaker as shown:

Reagent	36-cm Run	48-cm Run
Urea	18 g	28.8 g
40% (19 : 1) Acrylamide stock	5.3 mL	8.5 mL
dH_2O	25 mL	35 mL
Mixed-bed ion-exchange resin	0.5 g	1 g

Stir the solution until all the urea crystals have dissolved, and pass through a 0.2-μm cellulose nitrate filter, degas for 2–5 min, and transfer to a 100-mL graduated cylinder. The degas time should be constant for all gels, to ensure a reproducible polymerization rate for all runs (*see* **Note 7**).

Add filtered 10X TBE buffer and adjust the volume with dH_20 as follows. Never add the TBE before removing the mixed-bed ion-exchange resin, as this destroys the effectiveness of the buffer.

Reagent	36-cm Run	48-cm Run
10X TBE buffer	5 mL	8 mL
dH_2O	q.s. to 50 mL	q.s. to 80 mL

Polymerization begins when APS and TEMED are added to the gel solution, as described in the next subheading. You must work quickly at this point, so ensure the plates are mounted in the gray cassette, and that you have a clean, dry comb and 60-mL syringe close at hand.

3.3.4. Casting the Gel

Release the top clamp assembly and remove the comb. Add freshly made 10% APS solution (250 μL for a 36-cm run; 400 μL for a 48-cm run) to the gel solution, being careful not to introduce air bubbles. Swirl gently to mix. Add TEMED (35 μL for a 36-cm run; 55 μL for a 48-cm run). Swirl gently and thoroughly to mix and initiate polymerization.

Draw the prepared gel solution into a 60-mL syringe, filling it slowly to avoid introducing air bubbles. Ensure that the solution is at the tip of the syringe and that it contains no bubbles that could be injected between the glass plates. Insert the tip of the syringe into the luer fitting of the injection device. Gently depress the plunger of the syringe, allowing the gel mixture to flow between the plates. While maintaining a gentle pressure on the plunger, watch the front of the gel move toward the top of the plates, gently pressing or tapping the top plate if bubbles begin to form. Stop injecting when the solution reaches the upper notch region and completely fills all the space between the plates. Overfill slightly, so excess liquid will pool into the notched region. The "ears" of the plates must be filled with gel solution or a buffer leak can occur. With the syringe still attached, insert the straight edge of the shark's-tooth comb and verify that no bubbles are trapped at the comb–gel solution interface. The comb forms a single slot into which you later insert the teeth of the comb. Resecure the top clamp assembly against the plates. Make sure the top clamp is seated properly over the ears of the glass plates. If the comb is too loose, replace the top clamp with three or four large binder clips. Leave the bottom fixture in place and let the gel polymerize for at least 2 h (*see* **Note 8**). Check that polymerization has been successful by keeping an aliquot of the gel mixture (*see* **Note 7**).

Improved gel formulations (given in the following list) that increase read lengths have also been evaluated. Read lengths are longer with these formulations when compared to read lengths from 19:1 polyacrylamide gels ***(16)***. Longer sequencing read lengths can also be achieved using a revised buffer and gel system ***(17)***.

3.4. Electrophoresis and Data Collection

3.4.1. Instrument Setup

Before proceeding with this subheading, you should already have completed the appropriate sequencing reactions and prepared a polyacrylamide gel. While the gel is polymerizing, you can set up

Plate size and run speed	Gel formulations	Expected read length ranges
36-cm WTR plates; 2X (1200 scans/h)	4.5% 29:1 (acrylamide–*bis*-acrylamide) 4.8% PAGE-PLUS 5% Long Ranger	650–800 bp
36-cm WTR plates 4X (2400 scans/h)	4.5% 29:1 (acrylamide–*bis*-acrylamide)	550–700 bp
48-cm WTR plates 1200 scans/h	4.25% 29:1 (acrylamide–*bis*-acrylamide) 4.75% Long Ranger 5.25% PAGE-PLUS	750–900 bp

the sample sheet and run file for the data collection software (*see* **Subheading 3.4.2.**).

To prepare the plates for insertion into the 377, remove the injection device fixture and close the cassette clamps to secure the bottom of the plates. Flush the injection device with water to clean it. Remove the gasket, if necessary, to clean it. Clean the syringe in a similar manner if it is to be reused. Remove the top clamp and carefully pull out the casting comb from between the plates. With a damp, lint-free tissue clear any acrylamide from the notch in the upper plate. Remove small pieces of acrylamide from the well region to ensure that none fall into the slot when the shark's-tooth comb is inserted. Clean the glass plates with a lint-free tissue and dH_2O, particularly in the read region (approx 2.5–4.5 cm above the bottom of the glass; *see* **Fig. 1**) by wiping the plates in a single direction. Remove any salt deposits on the plates to ensure that they lie flat against the heat-transfer plates, when the cassette is loaded in the instrument. Clean the gel casting comb with dH_2O and allow it to air-dry so it can be inserted to form the wells for loading. Ensure the plates are free of dust, lint, or water spots, particularly in the read region. Turn the bottom set of plate clamps so the undercut part faces toward the top of the cassette. Move the beam stop into the down position and turn the plate clamps to hold it.

Carefully align the center registration line on the comb with the registration mark in the notch of the glass plates and slide the comb

between the plates, until all the teeth slightly depress the gel surface. If the gel surface is not completely flat, you may have to insert some of the teeth below the surface of the gel. If a tooth has penetrated the gel surface, do not attempt to withdraw the comb or samples may leak to adjacent wells. Place the upper buffer chamber against the top of the glass plates, making sure the overhang lip on the buffer chamber rests on the top of the glass plates. Close the top plate clamps to clamp it into position. The gel cassette is now ready to be inserted into the instrument.

If you have not already done so, prepare 1.4 L of 1X TBE buffer, as described in **Subheading 3.3.1.** To mount the gel cassette in the 377, open the door to the instrument and make certain the front and rear heat-transfer plates and the two small positioning pins at the bottom of the chamber are clean and dry (*see* **Fig. 2**). Place the lower buffer chamber in the bottom shelf of the electrophoresis chamber. Open the four cassette clamps in the electrophoresis chamber and place the gel cassette onto the cassette clamps and against the rear heat-transfer plate. The side spacers on the gel assembly that are located in the region of the rear plate notches should press against the two small positioning pins at the bottom of the chamber and the bottom of the plates should extend into the lower buffer chamber. Close the clamps to secure the cassette. Do not touch the gel plates within the read region.

Before loading samples or adding buffer, the gel is scanned to check for fluorescence on the plates or in the gel. If you opened a run window to create a run file while the gel was polymerizing (*see* **Subheading 3.4.2.**), display the window by choosing "Run" from the collection window menu. If you did not create a run file while the gel was polymerizing, choose "New" from the collection file menu, and click the "Sequence Run Folder" icon to open a "New Run" window. Choose "Plate Check" from the plate check popup menu and click "Plate Check;" the laser scans the plates without electrophoresis. The scan window that appears should show a relatively flat line across the screen in each of the four colors. If the scan line is flat, then the plates are clean. Proceed to prerunning the

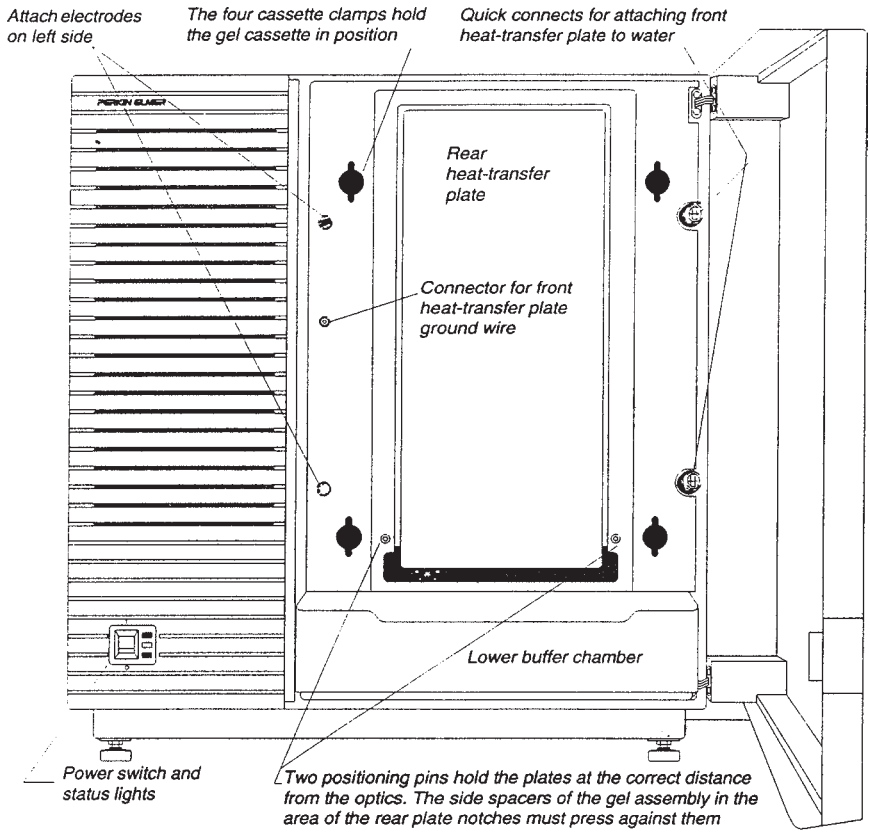


Fig. 2. ABI PRISM 377 electrophoresis chamber.

gel (*see* **Subheading 3.4.3.**). If peaks are seen in the scan window, click "Pause," remove and clean the glass plates in the read region, reinsert the plates, and click "Resume."

Before installing the front heat plate, fill the upper buffer chamber with approx 600 mL of 1X TBE and check for leaks. If a leak is found, siphon off the buffer and clean the gasket area of the chamber. Also, clean the area of the plate that contacts the gasket (*see* **Note 9**). Reassemble as before. Fill the lower buffer chamber with 1X TBE (approx 600 mL) until the bottoms of the glass plates are

covered. Avoid splashing. Make sure the lid is in place in the upper buffer chamber before starting the run.

To attach the front heat-transfer plate for 36-cm WTR runs, open the plate clamps between the upper buffer chamber and the beam stop. Push the front heat-transfer plate against the front of the cassette so it rests on the platform that is part of the beam stop bar. Secure it with the plate clamps and attach the water lines and ground cable. (**Note:** 48-cm runs do not require the front heat plate.) Plug in the electrode cables. The software should now be set up in preparation for loading samples.

3.4.2. Setting Up the Software

Before starting electrophoresis, the collection preferences need to be checked, and changed if necessary. Default parameters should be verified, or specified (via the preferences submenu from the 377 collection windows menu); this is described in detail in the 377 users manual. Default run parameters (e.g., modules, mobility files) are user specified dependent on the sequencing chemistry being used. Once the preferences have been set, the sample sheet can be created.

To launch data collection, choose "Restart" from the special menu. The Macintosh should be configured to launch 377 collection at startup. Before starting the run, sample information is recorded in the sample sheet; this information is copied into the run file which associates sample information with each lane position in the gel.

To create a sample sheet, choose "New" from the file menu, and click the sequence sample sheet icon. Fill in the spread sheet as appropriate—the minimum entry being a name or number for each sample loaded. Although the overall sample sheet settings for dye set/primer and matrix are determined by the preferences, the settings can be changed for individual samples. These changes are made from the popup menu activated by the triangle sign in the corner of the relevant cell in the document. When complete, save the sample sheet. From the "File-New" option, select the "Sequence Run Folder" icon to open a "New Run" window. Choose modules

for the plate check, the prerun, and the run in the module popup menus. For dRhodamine terminators, BigDye Primers, and BigDye Terminators the following modules should be specified:

Configuration	Run module	Base caller setting
2X (100 bp/h), 36-cm WTR	Seq Run 36E-1200	ABI100
4X (200 bp/h), 36-cm WTR	Seq Run 36E-2400	ABI200
48-cm WTR	Seq Run 48E-1200	Semiadaptive

Any prerun and plate check module will work on the ABI PRISM 377 DNA sequencer (including XL and 96 lane upgrades).

Use the correct dye set/primer (mobility) file as shown:

Sequencing chemistry	Mobility file
dRhodamine Terminator	DT (dR Set Any-Primer)
BigDye Terminator	DT (BD Set-Any Primer)
BigDye Primer	DP 5% LR (BD M13 FWD & REV)

Choose the sample sheet from the sample sheet popup menu. Select the type of comb, scan mode, separation distance, and matrix file from the respective popup menus. Check on "Autoanalyze with" if needed. Autoprint is best left unchecked since it is always useful to review the data before committing the files to print. The instrument is now ready for preelectrophoresis and sample loading.

3.4.3. Preparing and Loading Samples

To prerun the gel, flush all of the wells with 1X TBE (running buffer) and remove any bubbles under the bottom edge of the gel plates in the lower buffer chamber. Click the "Prerun" button and proceed to prepare your samples while this is in progress.

Resuspend your presequenced samples with loading buffer (5:1 deionized formamide: 25 m*M* EDTA with 50 mg/mL of dextran blue) as follows:

Comb (no. of wells)	36[a]	48[b]	64[b]	96[c]
Resuspension volume (μL)	4–6	4–6	4–6	4-6
Loading volume (μL)	0.75–2	0.75–1.5	0.75–1.5	0.5–1

[a]Full scan; [b]XL scan; [c]96 Lane scan.

Vortex-mix then centrifuge briefly. Heat samples at 95°C for 2 min to denature, then chill on ice. Vortex-mix and centrifuge the samples again, and place on ice until ready to load.

To load the samples, pause the "Prerun" from the run window, and carefully flush all the wells with running buffer. Immediately load sample into each odd-numbered well, and click "Resume" in the run window. Electrophorese for 2 min and click "Pause." Flush all of the wells with running buffer to remove any residual formamide from the previously loaded wells (*see* **Note 10**). Immediately load the even-numbered wells and click "Cancel" in the run window. Click "Run" to start data collection.

If you are loading 64 or 96 samples onto a gel, we recommend using an eight-channel pipet. With 96-lane gels, load samples into the bottom of the well close to the gel surface. For improved 96-lane gel resolution use the Sakabe/water run in protocol ***(18)***.

3.4.4. Observing Instrument Status and Data

During a run, instrument status can be checked to ensure that electrophoresis and laser scanning is taking place. To observe run parameters, choose "Status" from the window menu. This allows you to view the set electrophoresis values. The actual power and current readings during electrophoresis differ from the set values and change gradually throughout the run.

A log file is created for each run and stored in the run folder. To view the log file, choose "Log" from the window menu. The log file contains a comprehensive record of all error and status messages generated during a sequencing run.

You can view data from a run in real time during data collection. The scan window and the gel window display the data being collected. The real time data display differs from the display after analysis. On real time displays, the data collection program shows light intensities, color coded according to wavelength. Blue, green, yellow, and red (in that order) represent the wavelengths of the dye emissions within each dye set. Blue represents the shortest wavelength, and red represents the longest. These colors therefore repre-

sent the wavelengths of the dyes detected, rather than the bases being detected. For a fuller explanation please refer to the 377 user manual. The analysis program converts the information collected so that after analysis the colours are the same for each base. To view data in the scan and gel windows, choose "Scan" or "Gel Image" from the window menu respectively.

3.4.5. Cleaning Up After a Run

After the run is complete, remove the gel by first disconnecting the electrode leads, disconnect the front heat plate if used, and siphon off the buffer from the upper and lower chambers. Clean up any remaining liquid and remove the upper buffer chamber. Wipe the front and rear heat-transfer plates and the instrument positioning pins with damp lint-free tissue and dry them. Remove the gel plates from the cassette; wipe the cassette clean and dry it.

Pry the plates apart and remove the comb and spacers, lay a large Kimwipe on the gel and roll it up, lifting the gel of the plate. Rinse the plates with dH_2O. Rub the plates with a gloved hand to dislodge remaining gel pieces and rinse again. Rinse the buffer chambers, comb, and spacers with dH_2O.

3.5. Data Analysis

3.5.1. Working with the Gel File

At the conclusion of data collection, a run folder is created containing a gel file. The gel file stores the raw data collected during the run and a gel image. If the run was successful and the analysis settings correct, the gel file should be properly tracked. All that needs to be done at this point is to review the sample file data. If, however, the lane tracking is incorrect because the signal is weak and the tracker missed a lane, you may need to make some changes and then regenerate sample files. After automatic analysis, the gel file window opens and displays the newly created gel file. To open the gel file manually, double click the gel file icon.

It is important that the tracker lines are positioned directly over areas of data that display the strongest fluorescent signal. To review tracker line placement, in the gel file window, click the diamond lane marker for the tracker line you want to move, and use the keyboard left (or right) arrow keys to move the entire line left (or right). Individual nodes along selected tracker lanes can also be selected and moved in a similar manner.

Each time the tracker line positioning in a lane is changed, the sample data must be extracted to incorporate the new information into the sample files. To extract sample data, choose extract lanes from the gel menu and select the settings you want from the dialogue box that appears. To retain any tracking modifications, ensure that the gel file is saved before closing.

The ABI PRISM lane guide lane identification kit can be used to aid lane assignment and identification ***(19)***. Data generated using the lane guide kit can be collected using data collection version 2.6. Data are then analyzed using sequencing analysis software version 3.4.

3.5.2. Processing Sample Files

There are several reasons to reprocess a sample file or group of files after automatic file processing. Some examples include:

1. To correct initial setup errors (e.g., wrong instrument or dye set primer [mobility] file specified in the sample sheet).
2. To change the point where the software starts calling bases.
3. To use a different base caller or change the spacing estimate to improve the analysis results.

Files for analysis are added to the sample manager by choosing add files from the manager menu. Locate and open the folder that contains the files you want to add. To change a parameter value for files listed in the sample manager:

1. If the field has a checkbox, click to select and deselect.
2. If the field has a popup menu, open and highlight the value you want to select.

3. If the field has neither a checkbox nor a menu icon, double click the field to activate the text entry cursor, then type in the new value.

Files with the A box checked will then be analyzed when the start button is selected. If file analysis fails you can:

1. Change one or more parameter in the sample manager and reanalyze the affected sample file.
2. Check the sequence analysis error log for information about problems that occurred during analysis.
3. Check the run conditions to see if any problems occurred during data collection.

If you reanalyze a sample file, the previously analyzed results are overwritten. To avoid erasing the previous analysis results, save a copy of the sample file under a different name before you do the second analysis.

For a full description of the sequence analysis software and details of viewing and editing sample files, see the DNA Sequence Analysis User Manual.

3.6. Routine 377 Maintenance

To ensure optimal performance and reliability, certain routine maintenance tasks need to be performed. Problems in operation (*see* Troubleshooting, **Subheading 3.7.**) can usually be prevented, or solved, by completing these regular tasks.

1. Ensure all reagents are stored at the correct temperature to maximize their shelf-lives and do not use reagents that have exceeded their useful life-spans.
2. Be sure to clean the glass plates, heat-transfer plates, spacers, and combs regularly. Ensure the heat-transfer plate is stored safely when not in use, as surface damage to the plate may cause electrical arcing during electrophoresis. The gel cassette and the positioning pins in the electrophoresis chamber should be cleaned between runs; this is extremely important to ensure accurate focusing of the laser in the center of the gel.
3. The water reservoir located in the compartment on the right side of the instrument should be checked weekly. Refill the reservoir with

5% (v/v) ethylene glycol in deionized water, when it is between one third and one half full. To refill the reservoir, first ensure that the pump is not running, then unscrew the plastic reservoir and remove it by pulling downward, fill the reservoir to about three-fourths capacity and then replace the reservoir.

4. The Macintosh computer attached to the 377 requires regular attention and maintenance to operate efficiently and consistently. The Macintosh should be restarted before each 377 run to clear the random access memory (RAM) and provide sufficient memory to launch data collection. Unnecessary files should be removed from the hard drive at regular intervals, and the desktop file should be rebuilt monthly and after any new software is installed. To rebuild the desktop, hold down the apple and option (alt) keys while the Macintosh restarts. Continue to hold the keys down until the computer prompts you to rebuild the desktop file, click OK. A hard disk maintenance program such as disk first aid or Norton Utilities should be run at regular intervals. If Norton Utilities is installed on your Macintosh be sure to uncheck Norton filesaver and Norton Crashguard in extensions manager. If your computer is networked, it should be taken off during data collection, as problems can occur owing to network file backup programs that may be running. Keep SAM loaded on your hard drive and use it to inspect the hard disk routinely. SAM intercept should be unchecked in the extensions manager. Prevent program conflicts before they occur by using discretion when adding software programs. (Use the Macintosh only for ABI PRISM 377 software.)

3.7. Troubleshooting

1. If the firmware memory on the 377 becomes corrupted, it may be necessary to clear the instrument's memory so that a new firmware image can be copied into memory. Possible symptoms that may indicate corrupted memory include lack of communication between the Macintosh and the 377 or flat or no scan lines appearing in the scan window during a run. If this happens, press the reset button on the back of the 377 once. If this fails to establish proper instrument operation, carry out a total reset as follows. Quit the data collection program and press the reset button on the back of the 377, wait for 2–3 s and press it again. Start the data collection program during which a progress window appears "sending firmware image to instrument."

2. If an error occurs during data collection, it appears in the run log file. The run log lists both Macintosh errors and instrument firmware errors. If problems occur during instrument operation, check the error log for recent error messages, errors that occur during analysis are found in the analysis error log. Common error messages and a list of remedial actions can be found in the appropriate software user manual.

4. Notes

1. Care must be taken during plate washing and gel preparation to minimize residual background fluorescence on the gel image. The quality of water tends to be the largest variable between laboratories and therefore tap water is to be avoided during plate cleaning as this could be a source of fluorescent contamination and/or cause a buildup of organic material on the plates which may lead to reduced sensitivity or resolution. All plate washing should be done with high-quality deionized water, with all rinses being done in a direction away from the read region. Fluorescent contamination that is not removed by standard wash procedures or a catastrophic loss of resolution (bands tilted and poorly resolved) can be caused by material bound to the plates. Resolution problems may also be accompanied by gel extruding from between the plates during electrophoresis. In these instances, it will be necessary to soak the plates for 5 min in 1 *M* NaOH followed by 1 *M* Nitric acid, ensure that the plates are rinsed thoroughly with dH_2O. However, cleaning glass plates in a laboratory dishwasher with a hot deionized water rinse cycle is the most effective means of consistently and thoroughly removing surface contaminants ***(20)***. These procedures can also help prevent other issues associated with laser light scatter (*see* **Note 8**).
2. With BigDye chemistries it is possible to perform reduced volume reactions, on some templates (PCR products and some plasmids). For half-volume reactions dilute the ready reaction mix 1:1 with 2.5X sequencing buffer (200 m*M* Tris-HCl, 5 m*M* $MgCl_2$; pH 9.0 at room temperature), that is, 4 μL of ready reaction mix and 4 μL of 2.5X buffer for a 20-μL sequencing reaction. This ensures that the final buffer composition in a 20-μL reaction is 80 m*M* Tris-HCl, 2 m*M* $MgCl_2$; pH 9.0. The dilution buffer supplied by Applied Biosystems (P/N 4305605) is a 5X stock and should be diluted 1:1 with dH_2O to

give a 2.5X buffer. Template/primer constraints will govern how far the reaction premix can be diluted, for example, low purity or highly GC-rich templates may give poor or no signal at reduced volumes.

3. Characteristics of data from poor quality template preparations include top heavy data with a short length of read, weak signal strength, and overall noisy data. Typical contaminants in template preparations include salt and silica finings. It is possible to clean up a failed template preparation by centrifuging at high speed to pellet any precipitate, followed by purifying the DNA supernatant by ultrafiltration (e.g., with Centricon-100 microconcentrator columns). Alternatively extract the DNA supernatant twice with 1 vol of chloroform or chloroform–isoamyl alcohol in a 24:1 ratio. Precipitate DNA from the aqueous layer with ethanol to remove all traces of chloroform.
4. Removing dye primer sequencing reactions from under the oil (TC1 and 480 only) can be difficult because of their small (5 µL) volumes. If this is a problem, 5 µL of deionized water can be added to the sequenced products to give a total volume of 10 µL. If this is done, the volume of ethanol used for precipitation will need to be doubled. For BigDye Primers, an express load option that does not require ethanol precipitation can also be used. This is recommended for applications when template will be plentiful (e.g., PCR products and some plasmids). It is not recommended for difficult templates (e.g., BACs, PACs, or P1 clones) or for sequencing applications on the ABI PRISM 310 Genetic Analyzer. For express loading, combine the A, C, G, and T dye primer reactions with 5 µL of 5 m*M* EDTA, take 4 µL of this mixture, and add 4 µL of formamide/dextran blue loading buffer. Heat the open sample tube at 98°C for 5 min to reduce the volume and concentrate the sample. This can be loaded directly onto the 377. Aliquots in 5 m*M* EDTA will remain stable for at least 2 mo when stored at –20°C, and can be concentrated by ethanol precipitation if the express load option does not yield enough signal.
5. For sequencing, the polyacrylamide gel is one of the most important variables that determine the number of bases you can read. Impurities in any reagent (including dH_2O) can cause irreproducible gel porosity, deviant mobility, inhibition of polymerization, and poor resolution. Possible contaminants include acrylic acid, linear polyacrylamide, and ionic species (e.g., metals). For a full description of the polymerization process and factors that affect gel quality, *see* **ref. *21***.

6. **Safety note:** Urea causes eye, skin, and respiratory irritation. Acrylamide and *bis*-acrylamide are neurotoxins. Formamide is a known teratogen. Avoid inhalation and skin contact with these compounds and wear protective eyewear, gloves, and safety clothing. Always work in a fume hood when handling acrylamide solutions. Before using these or any other chemicals, always read the material safety data sheet for details on any potential hazards.
7. Using the gel recipes given, gel polymerization, once initiated with TEMED and APS, should take between 10–15 min. If polymerization takes much longer, this may indicate degraded APS or TEMED. Solid APS should look white and should "crackle" when water is added; once in solution it is unstable so needs to be made fresh. TEMED should look colorless and should be discarded if it turns yellow. The rate of polymerization is also affected by temperature and oxygen content. Polymerization at 20–23°C is optimal. Dissolved oxygen inhibits polymerization. To prevent problems caused by oxygen, strict attention must be paid to degassing gels. It is important to keep the vacuum strength and time constant during this step to ensure run-to-run reproducibility ***(21)***.
8. The term "red rain" has often been used to describe the pattern of vertical red streaks that are occasionally seen on sequencing runs. This is caused by laser light scatter from bubbles and/or channels which form in the read region as electrophoresis proceeds. This can be minimized or resolved by ensuring gel mixes are efficiently degassed. Plate washing with acid and alkali (*see* **Note 1**) can also remedy this problem. However, the most effective way to eliminate "red rain" is to seal the bottom of the polymerized gel with parafilm, or to pour the gel from the bottom with the gel injection device and leave the bottom fixture in place, until the run is set up. This ensures that the gel at the bottom of the plates does not dry out.
9. Electrical arcing is caused when buffer leaks between the front heat plate on the 377 and the glass plates. This can lead to damage to spacers/combs and pitting of the heat transfer plate. Salt buildup may also lead to this. We recommend that the heat plate is cleaned with dH_2O between runs to remove salt buildup. Also, make sure the gasket in the upper buffer chamber is not worn and is making a tight seal so that no buffer leakage occurs. Buffer leakage during a run will also cause electrophoresis failure, giving an instrument error alert, "no EP voltage detected." This alert can also be caused by depletion

of buffering ions. In this instance, it may be necessary to prepare new TBE buffer.

10. When you are loading many samples using a shark's-tooth comb and need to run them side by side, it is important that you load samples in alternate wells and electrophorese the gel briefly. Then rinse all the wells with 1X TBE and load the rest of the samples in the remaining wells. Formamide in the loading buffer helps narrow the lane width, but unloaded adjacent wells are required. Loading samples in adjacent lanes can blur the definition between the lanes. The automatic lane tracker in the analysis software needs to have discrete spaces between the samples to identify the lanes properly. Always place formamide loading buffer in the well to the left of your first sample lane, and in the well to the right of your last sample lane. The formamide in the buffer helps focus the bands in the first and last lanes. To ease sample loading, the wells can be visualized easier by preelectrophoresing a small amount of dextran blue/EDTA across the entire width of the comb. Rinse the wells prior to sample loading and the blue coloration that is left clearly marks the bottom of the wells.

References

1. Lee, L. G., Spurgeon, S. L., Heiner, C. R., Benson, S. C., Rosenblum, B. B., Menchen, S. M., Graham, R. J., Constantinescu, A., Upadhya, K. G., and Cassel, J. M. (1997) New energy transfer dyes for DNA sequencing. *Nucleic Acids Res*. **25,** 2816–2822.
2. Rosenblum, B.B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R., and Chen, S. M. (1997) New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res*. **25,** 4500–4504.
3. Human disease research (1996) *Biosystems Reporter* **224,** vol. 2; Applied Biosystems.
4. Disease gene discoveries made with ABI PRISM Systems (1997) Applied Biosystems.
5. Biotech breakthroughs (1998) *Biosystems Reporter* **29,** vol. 1; Applied Biosystems.
6. dRhodamine terminator cycle sequencing kit (1997) Product bulletin; Applied Biosystems.
7. Sequencing large DNA templates (1998) User Bulletin; Applied Biosystems.

8. Devine, S. E. and Boeke, J. D. (1994) Efficient integration of artificial transposons into plasmid targets in vitro: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res*. **22,** 3765–3772.
9. Taylor, R. G., Walker, D. C., and McInnes R.R. (1993) *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res*. **21,** 1677–1678.
10. The Qiagen guide to template purification and DNA sequencing, 2nd ed. (1998), Qiagen.
11. Optimizing PCR with AmpliTaq Gold DNA polymerase (1997) Application Note; Applied Biosystems.
12. The benefits of AmpliTaq Gold DNA polymerase in generating PCR templates for sequencing (1997) Application Note; Applied Biosystems.
13. Comparative PCR sequencing: A guide to sequencing-based mutation detection (1995) Applied Biosystems.
14. Precipitation methods to remove residual dye terminators from sequencing reactions. (1998) User Bulletin; Applied Biosystems.
15. Gel pouring with the redesigned cassette and fixtures (1998) User Bulletin; Applied Biosystems.
16. Improved gel formulations for extended sequencing read lengths (1997) User Bulletin; Applied Biosystems.
17. Achieving longer high-accuracy reads on the 377 sequencer (1999) User Bulletin; Applied Biosystems.
18. Optimizing resolution on the 96 lane upgrade (1999) User Bulletin; Applied Biosystems.
19. ABI PRISM lane guide lane identification kit (1999) Protocol; Applied Biosystems.
20. Cleaning glass plates to eliminate temporary loss of signal (1998) User Bulletin; Applied Biosystems.
21. Acrylamide polymerization—A practical approach (1997) Bio-Rad Laboratories, Inc.

13

Fluorescent Sequencing for Heterozygote Mutation Detection

Colin A. Graham and Alison J. M. Hill

1. Introduction

1.1. Direct Fluorescent Sequencing

Direct sequencing of PCR products using the dideoxy chain termination procedure developed by Sanger et al. ***(1)*** is now the most commonly used method for defining specific mutations. The main benefits of this method lie in its ease of use and this has been enhanced in recent years by the introduction of fluorescent labels and automated detection systems which obviate the need for radioactivity. Although the initial purchase price for automated sequencers is high, this is compensated for by single-tube reaction chemistry and rapid analysis and base calling.

1.2. Instrumentation

At the present time, three commercial fluorescent sequencers are in common usage throughout the world: the 373A/377/310/3700 fluorescent fragment analyzers (Applied Biosystems), the ALF and MegaBACE DNA sequencers (Amersham Pharmacia Biotech), and the LI-COR (MWG Biotech). The Applied Biosystems instruments

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

are the most widely used and benefit from the use of multicolor fluorescence detection. This enables single-lane sequencing and the company have developed very simple fluorescence dye-terminator sequencing kits. The methods described in this chapter are confined to the use of the 373A sequencer.

The 377 (Chapter 10) and 310 (Chapter 11) instruments are covered earlier in this book. The 373A instrument contains an electrophoresis chamber for vertical polyacrylamide gel electrophoresis. This enables single-base resolution of sequencing products. Fluorescently-labeled fragments pass through a "read" window, 24 cm from the loading wells, which is scanned by an argon laser. The fluorochromes are excited by the laser emission and are detected with filter wheels and a photomultiplier. The fluorescent signals are passed to an Apple Macintosh computer which analyzes their position and strength and produces a chromatogram consisting of colored peaks. The area under the peak represents the strength of the signal and the peak color is specific for the base at that position. The software gives a base call A, T, C, or G at each position or assigns N if the position is unclear.

1.3. Comparison of Sequencing Protocols

The methods and examples given in this chapter concentrate on the *Taq* dye terminator cycle sequencing system (**Fig. 1**), as this is now the most widely reported fluorescence sequencing method for the definition of mutations in human genes. Some of the methodological details to be considered before setting up a sequencing procedure for diagnostic use are now considered, bearing in mind that the method should be reliable, consistent, give good quality results, and be easy to use for a wide range of different sequences.

1.3.1. Sequencing Methods and Enzymes

Cycle sequencing using a thermostable DNA polymerase such as *Taq* (*Thermus aquaticus*) is the method of choice. The benefits of this method are: a large number of reactions can be performed simultaneously using a thermal cycler, only a small amount (<1 μg)

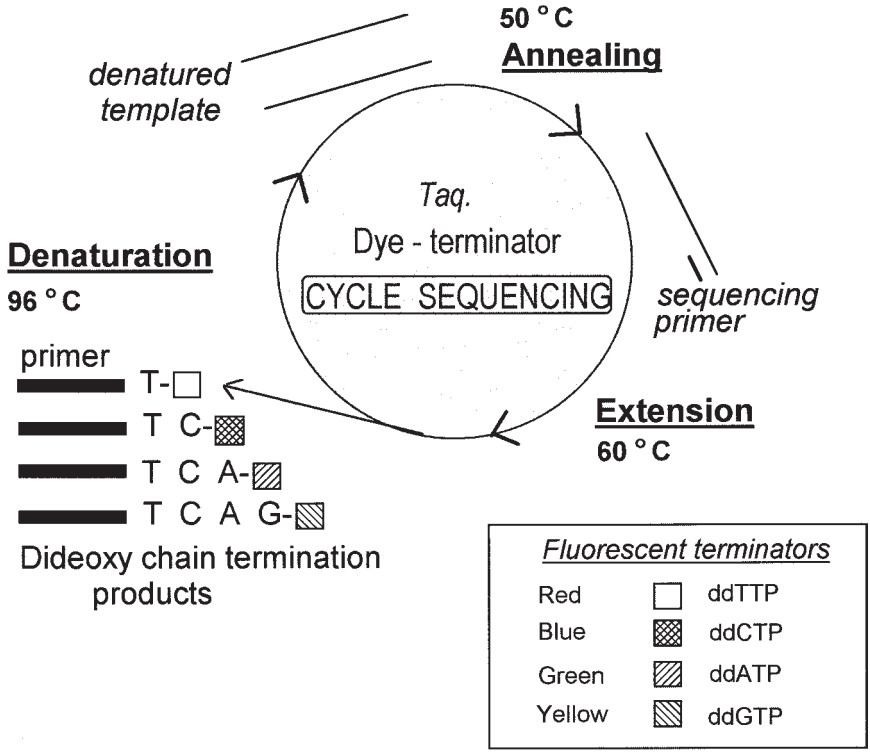


Fig. 1. *Taq* dye terminator cycle sequencing. Cycle sequencing is like a linear PCR reaction, using single- or double-stranded DNA template and priming from a single sequencing primer. The dideoxy chain terminators carry four distinct fluorescent dyes; thus, all the termination products can be loaded into a single lane and size separated on a denaturing polyacrylamide gel. The sequence is read as a color code using a laser and fluorescence detection system.

of template DNA is required to give good quality sequence, and the cycling nature of the reaction gives strong signal profiles and greatly reduces strand annealing and random priming events that produce background noise.

The use of *Taq* polymerase enables sequencing to be carried out at 60°C, and this reduces secondary structure in the template and gives more efficient read through GC-rich regions. One of the

disadvantages of *Taq* is that incorporation efficiencies for the fluorescent terminators are sequence dependent and this can lead to uneven signal profiles. This problem has been resolved to some extent by the introduction of a modified *Taq*–Fluorescent Sequencing (FS) polymerase that has essentially no 5' > 3' nuclease activity and also has a greatly reduced discrimination for dideoxynucleotides, thus permitting much lower concentrations of dye terminators to be used in the sequencing reaction. This results in improved terminator incorporation efficiency, more even peak heights, and a simpler protocol for dye terminator removal. The introduction of dRhodamine and BigDye (energy transfer dyes) terminators have further improved the uniformity of peak heights and reduced noise levels; however, these chemistries cannot be used on the 373A if it is to be used for the detection of fluorescent polymerase chain reaction (PCR) fragments using GeneScan software, as well as sequencing. In such cases, the *Taq*–FS ABI PRISM dye terminator kit is the best option.

If sequence quality still remains a problem, then dye primer chemistry should be considered. In dye primer sequencing, the fluorescent labels have to be attached to the sequencing primer and four separate reactions are required matching a specific dye-labeled primer with a specific dideoxy terminator. The products are then pooled and cleaned to remove dye primers prior to electrophoresis. With dye terminator sequencing, the dye labels are attached to the dideoxy terminator molecules and thus sequencing can be accomplished in a single tube reaction. Technicalities of the different sequencing chemistries are discussed by Hawkins et al. ***(2)***.

The cost and ease of use benefits of the dye terminator system make it the method of choice for the majority of PCR product sequencing requirements. If repeated sequencing of the same region is required, then the dye primer method should be considered.

1.4. Primer Design and Synthesis

Primer base composition, secondary structure, stability, and specificity are all important in producing clear sequence data. The

primer sequence should be specific to reduce binding to secondary sites and a minimum length of 20 base pairs is recommended. Primers with a high G + C content have higher annealing temperatures and will be more stable and effective in cycle sequencing. The presence of a CG basepair at the 3' end of a primer can help to stabilize it during the cycling reaction. Computer packages, such as Primer Detective (Clonetech), can be used to identify suitable primers and will show any self- or cross-complementarity in the selected primers. Primer length, C + G content, and product melt temperature are also considered. It can still be difficult to design reliable primers for some regions, mainly because the DNA sequence surrounding the area of interest contains repeat sequences or is AT-rich. Poor quality primers can result in relatively good peak signal strengths but high background noise or noisy signal with no well-defined peaks; this can be because of random priming reactions if the primer is not specific or to impure primer. Primer synthesis for sequencing should be carefully monitored as short-mers and impurities may interfere with the sequencing reaction. For optimal results, sequencing primers can be high-performance liquid chromatography (HPLC) purified, but this is not generally required. If the primer, or template, concentration is too high, then reaction components can become exhausted during the cycle sequencing reaction with considerable reduction in signal intensity toward the end of longer fragments.

2. Methods

The methods described are specifically designed for performing fluorescent DNA sequence analysis on the Applied Biosystems Model 373A. Use of the 373A sequencer and the relevant software is described in detail in the manufacturer's manual. Related DNA sequencers such as the 377/310 are covered in Chapters 10 and 11, respectively.

2.1. Template Preparation

The methods described in this chapter deal exclusively with template DNA produced using the PCR. The yield and purity of the

amplified product are of critical importance for good quality sequence. Thus, 5 μL of the amplified DNA is run out in 2% agarose gels to assess the concentration and purity. If nonspecific product is seen, the amplification reaction should be repeated under more stringent conditions, in the presence of dimethyl sulfoxide (DMSO), at a higher annealing temperature or after band-stab gel purification as described in **Subheading 2.1.1.** ***(3)***.

2.1.1. Purification of PCR Amplified DNA by "Band-Stab"

This method of template purification allows amplification of specific regions of DNA without any nonspecific contaminating products, occasionally produced during the PCR reaction.

1. The section of DNA of interest is amplified by PCR and run out on a 2% agarose (Life Technologies) gel containing ethidium bromide (0.5 μg/mL).
2. The gel is visualized over a UV transilluminator and a sterile needle is inserted into the correct DNA band in the gel.
3. DNA picked up by the needle is redissolved in 50 μL of sterile deionized water in a clean microcentrifuge tube.
4. The tube is microfuged at maximum speed for 10 min.
5. Five microliters of the resulting supernatant is used as template in a second PCR reaction with the same primers (*see* **Note 1**).
6. An aliquot of the second PCR product is again checked for concentration and purity on an agarose gel.

2.1.2. Purification of Template DNA

The PCR product that is to be used as template for the sequencing reactions needs to be purified to remove any residual primers or excess PCR reaction components that could interfere with the sequencing reactions. The use of ultrafiltration spin columns Centricon 100 or Microcon YM-100 (Amicon Bioseparations-Millipore) has proved to be very effective. Alternative methods of template purification are given in **Note 2**.

2.1.2.1. Centricon-100 Purification Columns

1. Assemble Centricon-100 columns as described in the manufacturer's manual.
2. Add 2 mL of sterile, deionized water to the top of the column.
3. Gently layer the template DNA (30–40 μL of PCR product) on top of the water.
4. Attach the collection tube provided to the top of the assembled column.
5. Centrifuge at 700*g* for 30 min and then remove lower collection chamber and discard fluid (a second centrifugation with a further 2 mL of water is not required).
6. Invert the column and centrifuge for 2 min to collect the purified sample (~50 μL) .
7. Transfer the sample to a clean, sterile microcentrifuge tube and store at –20°C.

2.1.2.2. Microcon-YM100 Purification Columns

1. Assemble Microcon-100 columns as described in the manufacturer's manual.
2. Add 500 μL of sterile, distilled H_2O to the top of the column.
3. Gently layer the template DNA (30–40 μL of PCR product) onto the top of the water.
4. Centrifuge the column in a microfuge (MSE) for 1–2 min at 3500*g*.
5. Add 500 μL of sterile, distilled H_2O to the top of the column and centrifuge again for 1–2 min.
6. Add 30 μL of sterile, distilled H_2O to the column, invert, and centrifuge briefly to collect the purified sample.
7. Transfer to a clean, microcentrifuge tube and store at –20°C.

The purified samples can be checked on 2% agarose gels to ensure efficient sample recovery from the column.

2.2. Dye Terminator Cycle Sequencing of Double-Stranded PCR Fragments

The method described here relates to the single-tube sequencing of double-stranded DNA. Sequencing is carried out using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit with Ampli*Taq* DNA Polymerase, FS (Applied Biosystems).

2.2.1. Cycle Sequencing Reaction

Each of the sequencing kits contains reagents for 100 sequencing reactions (*see* **Note 3**). The primers used in the sequencing reactions may be the same ones used in the original PCR reactions or they may be specifically designed for sequencing and internal to the original primers. In either case, they should be diluted to a working concentration of 3–5 pmol/μL.

1. Set up reaction mix in a clean, sterile microcentrifuge tube.
   Reaction mix: $N \times 8.0$ μL of reaction premix
   $N \times 1.0$ μL of sequencing primer
   $N \times 6.0$ μL of sterile, deionized H_2O
   (where N = no. of samples + 1)
2. Add 5 μL of template DNA into each PCR reaction tube (200 or 600 μL depending on the PCR machine in use).
3. Aliquot 15 μL of reaction mix into each of the PCR tubes.
4. Overlay reactions with one drop of mineral oil, if necessary and place on the thermal cycler preset with the appropriate cycle sequencing reaction (*see* **Note 4**).

2.2.1.1. Typical Cycle Sequencing Reaction for 2400/9600/9700 (Oil-Free) (Applied Biosystems)

	96°C	10 s
25 cycles	50°C	5 s
	60°C	4 min
Final cooling	4°C	10 min

2.2.1.2. Typical Cycle Sequencing Reaction for 480 (Requires Oil Overlay) (Applied Biosystems)

	96°C	30 s
25 cycles	50°C	15 s
	60°C	4 min
Final cooling	4°C	10 min

2.3. Purification of Sequencing Reaction Products

After cycle sequencing, the excess fluorescent dye terminators used to drive the reaction must be efficiently removed for good quality electrophoretic separation of the termination products and effective analysis of the initial 50 bases after the primer. This is achieved by ethanol precipitation (*see* **Note 5**).

1. Transfer the 20 μL of sequence reaction to a clean 0.6-mL microcentrifuge tube, taking care not to transfer any of the mineral oil overlaying the reaction (when used).
2. Precipitate reactions with 3 μL of 2 *M* sodium acetate, pH 4.5 (Applied Biosystems) and 300 μL of 100% ethanol at +4°C.
3. Place samples on ice for 10–15 min.
4. Microcentrifuge at 14,000*g* for 30 min.
5. Wash the pellet with 70% ethanol, taking care not to dislodge the pellet. Microcentrifuge at 14,000*g* for 15 min.
6. Remove ethanol carefully and vacuum-dry the pellet.

2.4. Polyacrylamide Gel Preparation and Electrophoresis

The glass plates used for fluorescent sequencing do not contain any fluorescent elements and are obtained from Applied Biosystems. They should be free of dust and grease and are cleaned using the detergent Alcinox (Aldrich Chemical). After washing, the plates are rinsed extensively in tap water, deionized water, and finally 100% ethanol before air-drying. The plates are separated using 0.4-mm spacers and poured horizontally on a level surface. If vertical pouring is preferred, then the plates are taped along the outer edges.

2.4.1. Gel Casting

1. Mix 50 mL of Sequagel-6 (premixed 6% sequencing gel, 19:1 acrylamide:*bis*-acrylamide with TBE buffer and 6 *M* urea) (National Diagnostics) with 400 μL of fresh ammonium persulfate (10% solution), (Bio-Rad Laboratories) (*see* **Note 6**).
2. Set the plates on a level surface and add the gel mix from the top center using a 20-mL syringe, tapping the plates occasionally to ensure air bubbles are not trapped. Small air bubbles at the edges are not a problem.
3. Insert and clamp the well-former and clamp the top edges of the plates with bulldog clips.
4. Allow the gel to polymerize for 1–2 h. If storing overnight, wrap the top and bottom with clingfilm. Clean off any dust or polymerized polyacrylamide from the glass plates before installing in the 373A.
5. Install the plates into the 373A instrument and use the "Plate Check" function to check the plates for any dust or spilled acrylamide as

described in the User's Manual (set the blue line to 500–800 for optimal sensitivity). If necessary, clean the outside of the plates again with deionized water and 100% ethanol. Scratches on the plate or irregularities cast in the gel can be lifted above the scan window by placing cardboard under the lower buffer chamber.

6. Carefully remove the well-former and insert the shark's tooth comb with the teeth of the comb just touching the surface of the gel (24- and 36-well combs are available). Fill the buffer containers with 1X TBE (National Diagnostics).
7. Set up the 373A DNA sequencer according to the manufacturer's instructions and select "Prerun" function to prerun the gel at 30 W for up to 1 h.

2.4.2. Preparation of Samples for Gel Electrophoresis

Dried pellets may be resuspended in loading buffer and stored at 4°C for 1–2 h prior to gel loading.

1. Prepare fresh loading buffer: 5 μL of deionized formamide (Sigma-Aldrich) and 1 μL of 50 m*M* EDTA, pH 8.0 (diluted from 500 m*M* EDTA [Life Technologies]), per sample.
2. Dissolve the pellet in 4 μL of loading buffer and denature the samples at 90°C for 3 min and place on ice.
3. Abort the "Prerun" if it is not complete.
4. Wash out the wells with 1X TBE using a syringe with spade tip to remove urea, and load samples into the odd-numbered wells only. Run samples into the gel for 5 min. Interrupt the run and load the even-numbered samples. Staggered loading allows the sequence analysis software to more easily identify individual sample lanes for tracking purposes.
5. The electrophoresis setting are 25–30 W and 2500 V with the power limiting.
6. Click on data "Collect" on the computer screen data collection icon and to ensure data collection, check that the scan numbers start to increment as data is stored on the computer. If this does not occur, there is probably some problem with the 373–computer interface.

2.5. Analysis of Results

The sequencing software analyses a run by looking for a first fluorescent band that is followed within a given time period by a second

band. On this basis, the software designates the first band as base 1. It can, however, be confused by salt fronts running ahead of the main data or by the presence of unincorporated terminators that give rise to signal dye "blobs" and result in inappropriate base calling and reduction of the observed signal strength in the true sequence data. In addition, the first 20 or so base calls can be difficult to read because of high signal levels, and it is therefore important to be able to assign base 1 accurately and reanalyze that portion of the sequence giving the best data. This can be achieved by the postrun operator-controlled removal of nonspecific data before and after the readable sequence.

It is also important for the software to recognize each individual lane and to stay tracked within the center of the lane. Tracking can be adversely affected by uneven polymerization of the gel or large residual amounts of salt in the samples. The dye blobs associated with unincorporated dye terminators can also spill fluorescence into adjacent lanes and confuse tracking. It is therefore necessary to check that the assigned tracks remain in the center of the lane for the length of the run, and if not then they should be reassigned and the lanes reanalyzed. This type of operator-controlled reanalysis can significantly improve the data quality. The newest versions of the analysis software include the "neural net tracker," which has greatly improved tracking ability (*see* **Note 7**).

2.5.1. Signal Strength

As mentioned previously, the incorporation efficiency of fluorescent dye terminators by *Taq* polymerase is variable and the C terminator is the least efficient. Thus, the C signal level is usually the lowest unless the sequence is very C rich. If C is <50, there will be a large amount of C noise in the chromatogram profile. Such sequence profiles are best disregarded and the sequencing repeated with an increase in the amount of template.

2.5.2. Base Spacing and Base Calling

To accurately call the bases in a sequence, the sequencing software needs to determine average peak intensities and spacings over

a minimum of approx 150 base pairs. Fragments shorter than this result in the software not being able to accurately assign base spacing. When this happens, the software is more susceptible to effects of gel compressions which are thought to be due to the presence of hairpin loop structures at the ends of fragments. The mobility of shorter fragments is more affected by these secondary structures than longer fragments, and thus they are more likely to show compression effects. These manifest themselves as uneven spacing in the chromatograph and, in the worst instance, the computer software will insert a base into the gap that it assumes should be filled. If the background noise is high in the gap, then the software might call a base. In many instances, however, it will make an equivocal call (*N*). The base spacing can be called only if the scans per base fall between 9 and 15. This is determined by the rate of electrophoresis, which can be adjusted by altering the electrophoresis power for the sequencing run or optimizing the polymerization of the polyacrylamide gel. Base calling can also be compromised by high noise levels because of low signals or false priming.

2.5.3. Sequence Alignment and Comparison Software

The Sequence Navigator software (Applied Biosystems) program can perform multiple alignments efficiently and can consider peak size for heterozygote detection. A composite normal can be produced which can then be compared with abnormal samples for heterozygote detection and the level of recognition of the heterozygote can be altered by the user. The program can also highlight sequence ambiguities when comparing aligned sequences. When using 373 derived sequence files, the program can display multiple sequence chromatograms, and these can be aligned and scrolled simultaneously for visual comparison of sequences (this is achieved by holding the option key down while scrolling on any one sequence).

2.5.4. Data Storage

For each 373A instrument, it is advisable to have a Macintosh computer dedicated to collection of data and initial automated analy-

sis, which allows two to three runs per 24-h period with virtually continuous data collection. For data collection, it is recommended that there is at least 40 Mb of free hard disk space on the computer prior to collection and that the disk is regularly monitored and defragmented with a repair program such as Norton Utilities. To process this amount of data from the instrument, it is necessary to have two additional computers for detailed sequence analysis, comparison, and printing of hard copies. Networking of the computers to a central server unit is preferred as transfer of large files (up to 20 Mb) is required. This can otherwise be achieved by the use of removable hard or optical drives or Zip drives. Once analysis of the large gel file is complete, this file can be discarded and only the results files require archiving. A result file for a single track in a gel occupies 96 K of disk space; thus two floppy disks (1.4 Mb) are required to store result data from a 24-lane gel. It is recommended that bulk archiving is carried out using removable hard, optical, or CDR disks with capacities of approx 40–800 Mb. However, essential results should also be stored on a floppy disk. Networking to a mainframe computer for storage is another option that should be discussed with your local information technology department, as it is now one of the best solutions because servers with large capacity RAID drives are now very affordable. A tape-based backup system should still be employed for the server unit.

3. Fluorescent Sequencing for Mutation Detection

Taq dye terminator sequencing is now routinely used in many diagnostic laboratories for heterozygote mutation detection. This system provides good reliable results, providing that the protocols are well standardized and the deficiencies are recognized.

3.1. Technical Priorities for Fluorescent Sequencing

1. PCR product quality.
2. Sequence with Forward and Reverse primers.
3. Standardize the electrophoresis procedure (use a commercial gel mix).
4. Ensure that the thermal cycler is suitable, preferably oil free.

5. Use fresh sodium acetate to precipitate extension products.
6. Check that data are being collected at the start of the run.
7. Reanalysis can improve the quality of the sequence data.

Visual pattern recognition is very efficient for identifying sequence profile irregularities providing that the *Taq* dye terminator sequencing anomalies and the heterozygote detection criteria defined in **Subheading 3.2.** are considered. A color inkjet printer such as the Hewlett-Packard 500–800 series or the Epsom 700 series is more suitable for printing sequence chromatograms than the thermal wax printers previously supplied with the machine because of cost and the color definition is better.

3.2. Heterozygote Detection

1. Print sequence profiles in color at 4–5 panels per page and 500–800 points per panel.
2. Signal strength reduction of the normal base is the most consistent feature denoting a heterozygous base position (**Figs. 2–4**).
3. The appearance of the second base in a heterozygote is sometimes not detected by the base calling software (**Fig. 2**).
4. The base after the heterozygote position will often show an altered signal strength (**Fig. 2**).
5. Heterozygotes that show only peak reduction in one direction generally show the mutant base clearly in the other direction (**Fig. 3**).

Fig. 2. *(opposite page)* Heterozygote mutation detection in the α-*galactosidase A* gene in patients with Fabry's disease. Fabry's disease is an X-linked recessive disorder with affected males (row 2) showing as hemizygotes for the gene mutations and female carriers (row 3) showing as heterozygotes. The exon 5 R220X mutation shows the classical pattern with the normal C base changed to a T in the male mutant and the female carrier showing both bases, although the software fails to detect the heterozygote. The exon 6 *N298K* mutations shows how a change in the size of the base after the mutant base can often aid in the identification of mutations. The exon 7 4-bp deletion mutation shows how in the heterozygote the reading frame is staggered after the mutation.

α-galactosidase A

Exon 5

R220X

CGA - TGA

Exon 6

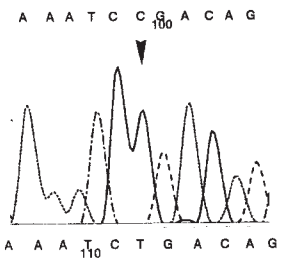
N298K

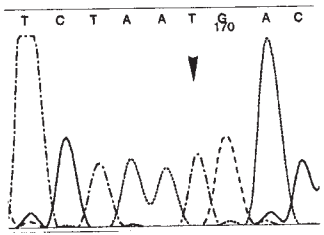
AAT - AAG

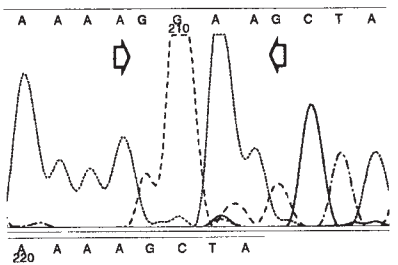
Exon 7

1174 del GGAA

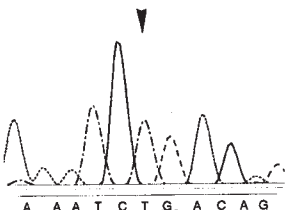
NOR

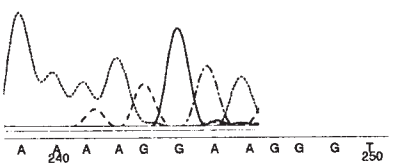




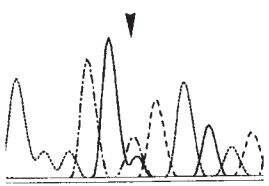


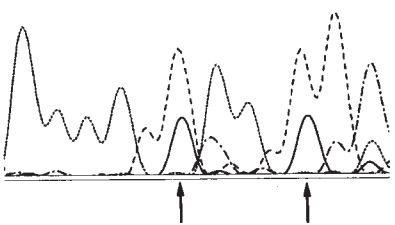
MUT





HET





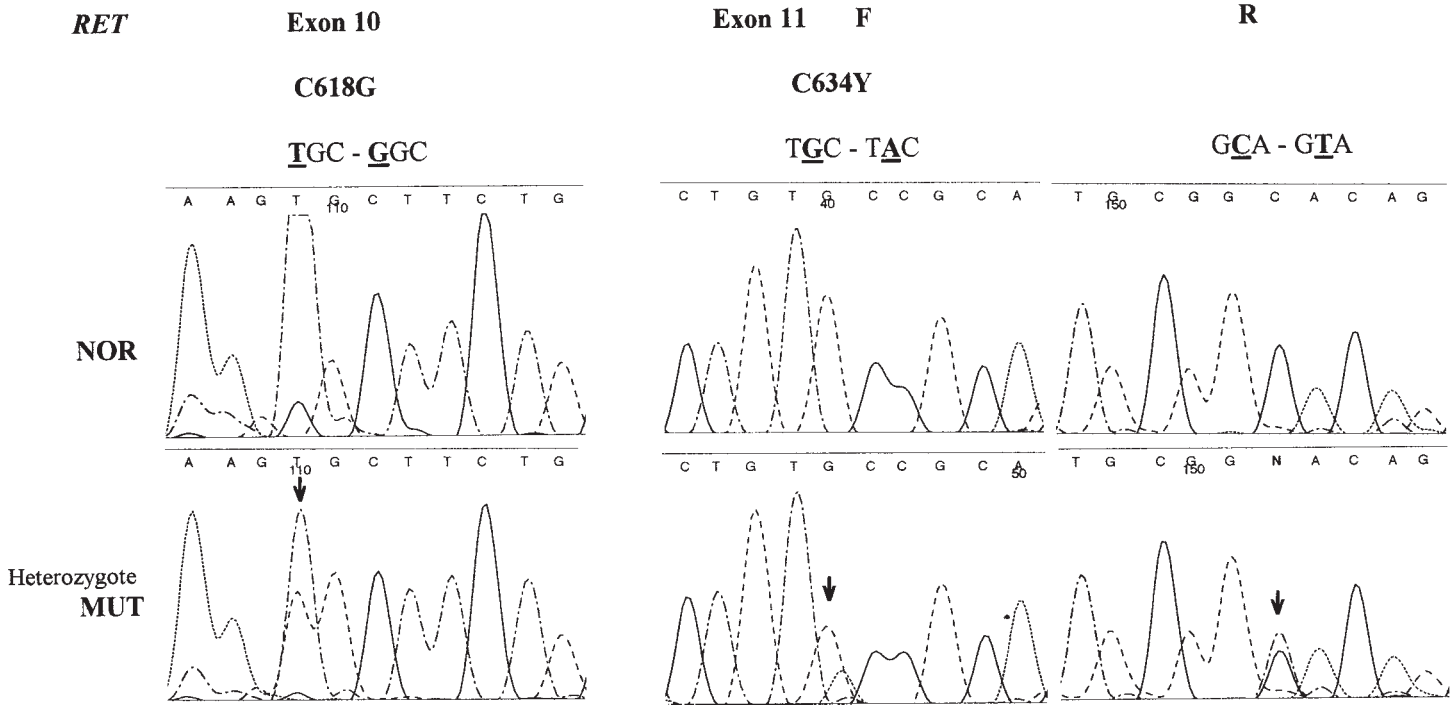


Fig. 3. Heterozygote mutation detection in the *RET* oncogene in inherited cancers. The exon 10 *C618Y* mutation shows a heterozygote germline mutant associated with thyroid cancer. The exon 11 mutation *C634Y* is the most common germline mutation causing MEN 2a. Here, the decrease in the signal strength of the G base is clear in the mutant, when sequenced in the forward direction; however, the mutant base is of low intensity, but in the reverse direction the mutant base predominates. Thus, bidirectional sequencing is very important for heterozygote detection.

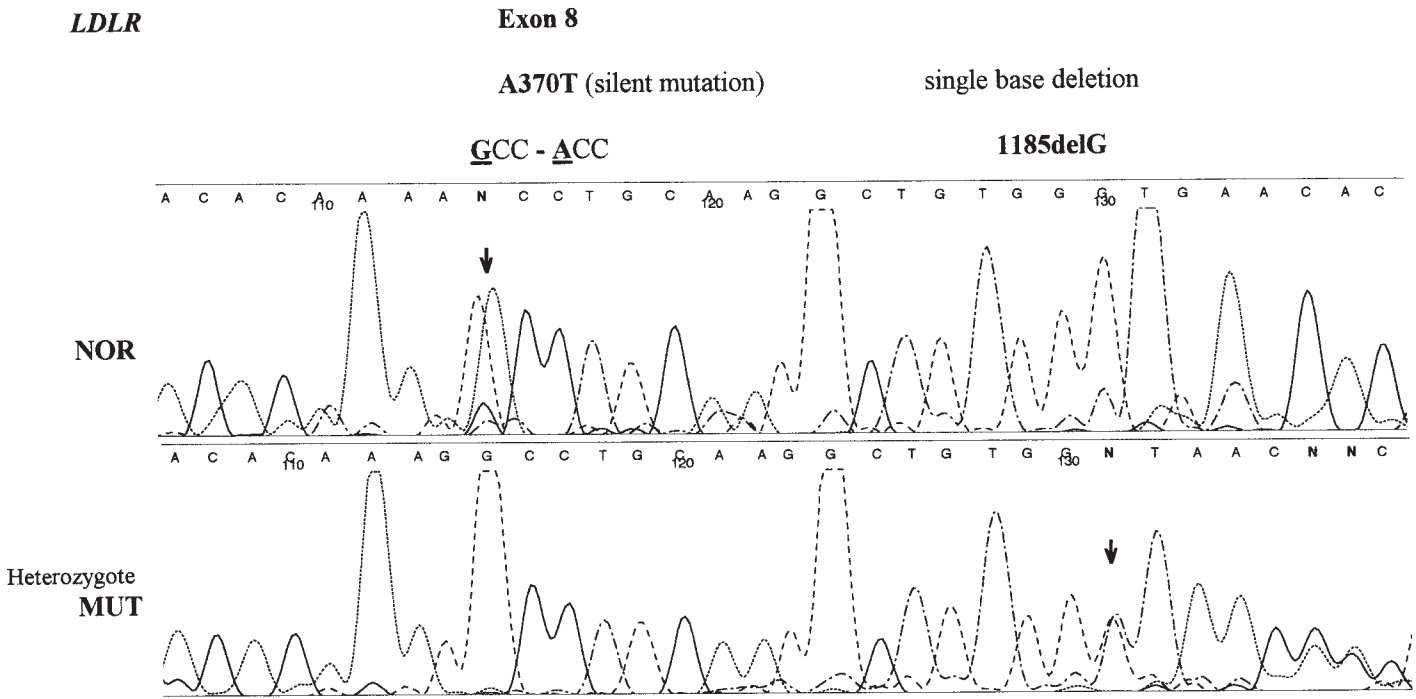


Fig. 4. Heterozygote mutation detection in the *LDLR* receptor gene in patients with heterozygous familial hypercholesterolemia. The exon 8 variant *A370T* shows as a classical heterozygote in the normal individual, as this variant is a silent mutation which does not cause disease. The heterozgote mutant (row 2) shows a single base deletion that shifts the reading frame back one base as indicated by the arrow.

3.3. Taq-*FS Dye Terminator Sequencing Anomalies*

1. Peak profiles are irregular but highly consistent for a given sequence.
2. A G base following an A base often has a greatly reduced signal.

3.4. Examples of Fluorescence Dye Terminator Sequencing for Heterozygote Mutation Detection

Some examples of the use of dye terminator fluorescent DNA sequencing for heterozygote detection as applied within a diagnostic molecular genetics laboratory are given in the following paragraph.

Mutation detection in the X-linked recessive disorder, Fabry's disease, which is a result of mutations in the α-*galactosidase A* gene ***(5)***, is shown in **Fig. 2**. Mutation detection is important for heterozygote carrier identification in the disorder, as serum levels of the enzyme are often equivocal in carriers.

Mutations analysis in the *RET* proto-oncogene in multiple endocrine neoplasias (MEN2a, 2b), thyroid cancer (FMTC), and Hirschsprung disease is important for accurate diagnosis of these conditions and for counselling and testing other family members ***(6)***. The germline mutations largely occur in cysteine residues in exons 10 and 11 of the *RET* gene as illustrated in **Fig. 3**.

Mutation detection in the low-density lipoprotein gene (*LDLR*) in heterozygous familial hypercholesterolemia is shown in **Fig. 4**. This illustrates that not all DNA base changes that alter the amino acid are disease causing as is the case for A370T. DNA analysis is important for presymptomatic diagnosis in this condition so that individuals who are heterozygous for a mutant allele can receive optimal patient management and treatment as early as possible to help prevent premature heart disease ***(7,8)***.

Automated fluorescent sequencing using the dye terminator chemistry is increasingly being used in diagnostic and research laboratories to define rare mutations detected by a rapid screening procedure such as denaturing gradient gel electrophoresis (DGGE) ***(9)***, single-stranded conformational polymorphism (SSCP) ***(10)***, and protein truncation testing (PTT) ***(11)***. The main benefits of this system are the single-tube and single-lane reaction chemistry, cycle

sequencing for easy sequencing of PCR products, fluorescent detection eliminating the need for radioactivity, automated base calling, and semiautomated analysis. The availability of hard copy printouts of sequence profiles and the ease of mass storage of data are also very important attributes of automated fluorescent sequencing.

4. Notes

1. Sometimes a 1:10 dilution of the band-stab DNA works better for reamplification.
2. Alternative methods can be used for template purification ***(4)*** including ion-exchange columns such as QIAquick-spin columns (Qiagen); "gene cleaning" methods using glass powder suspensions such as Geneclean (Stratech Scientific), or similar DNA binding agents as used in Magic PCR preps (Promega). These methods may be less expensive than the ultrafiltration method described earlier; however, they are more labor intensive, and we have not found them to give as consistently good results. Overall, the Centricon-100 procedure gives the most consistent results in our hands and is recommended for diagnostic sequencing. This method efficiently purifies PCR products of >100 bp.
3. The reaction mix can be diluted 1:1 with half TERM buffer (Genpak) without any apparent loss of sequence quality, thus virtually halving the cost of each reaction.
4. The cycling reaction should take approx 2 h and 45 min. If the thermal ramping time is too fast or too slow, the resulting data may be poor with high levels of background noise. The following thermal cyclers have been found to give good quality cycle sequencing Perkin-Elmer 480, 2400, and 9700.
5. The precipitation step is vital for efficient recovery of termination products after phenol–chlorofom extraction and it is necessary to use fresh sodium acetate (<1 mo at 4°C after opening) to obtain good product yield. It is also ***not*** recommended to resuspend the pellet during the 70% ethanol rinse, to remove the sodium acetate prior to electrophoresis, as this can result in a low yield and poor signal.
6. The use of a low fluorescence commercial sequencing gel mix is highly recommended, as these are batch tested and thus gel consistency can be assured within a batch. This also eliminates gel variation problems if the machine is being operated by multiple users.

The National Diagnostics gel mix does not require degassing for optimal polymerization.

7. Neural net tracker software is recommended when running gels with 64- or 96-well sample combs. One of the few drawbacks of this system is its relative inability to cope with blank or empty lanes, with the subsequent need for operator-controlled correction which can prove difficult with higher sample numbers. These versions of the software also require a G3 Macintosh computer (minimum) to operate efficiently and are available only for upgarded and stretched 373A sequencers, but not for the standard model.

References

1. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
2. Hawkins, T. L., Du, Z., Halloran, N. D., and Wilson, R. K. (1992) Fluorescence chemistries for automated primer-directed DNA sequencing. *Electrophoresis* **13,** 552–559.
3. Bjournson, A. J. and Copper, J. E. (1992) Band-stab PCR: a simple technique for the purification of individual PCR products. *Nucleic Acids Res*. **20,** 4675.
4. Rao, V. B. (1994) Direct sequencing of polymerase chain reaction-amplified DNA. *Anal. Biochem.* **216,** 1–14.
5. Davies. J., Winchester, B. G., and Malcolm, S. (1993) Mutation analysis in patients with the typical form of Anderson-Fabry disease. *Hum. Mol. Genet.* **2,** 1051–1053.
6. Donis-Keller, H., Dou, S., Chi, D., et al. (1993) Mutations in the *RET* proto-oncogene are associated with MEN 2a and FMTC. *Hum. Mol. Genet.* **2,** 851–856.
7. Hobbs, H. H., Browm, M. S., and Goldstein, J. L. (1992) Molecular Genetics of the LDL receptor gene in familial hypercholesterolaemia. *Hum. Mutat.* **1,** 445–466.
8. Graham, C. A., McLean, E., Ward, A. J., et al. (1999) Mutation screening and genotype:phenotype correlation in familial hypercholesterolaemia. *Atherosclerosis* **147/2,** 309–316.
9. Hughes, D. J., Hill, A. J. M., Macek, M., et al. (1996) Mutation characterisation of CFTR gene 206 Northern Irish CF families: Thirty mutations, including two novel, account for 94% of CF chromosomes. *Hum. Mutat.* **8,** 340–347.

10. Chillon, M., Casals, T., Gimenez, J., et al. (1994) Analysis of the CFTR gene in the Spanish population: SSCP screening for 60 known mutations and identification of 4 new mutations (Q30X, A120T, 1812-1 G>A and 3667del4) *Hum. Mutat.* **3,** 223–230.
11. van der Luijt, R., Khan, P. M., Vasen, H., et al. (1994) Rapid detection of translation-terminating mutations at the Adenomatous Polyposis Coli (*APC*) gene by direct protein truncation test. *Genomics* **20,** 1–4.

14

Sequence Databases and the Internet

Anne E. Hughes

1. Introduction

Every researcher who generates new sequence data will have a curiosity and a need to know if his or her sequence is the same or similar to one that has been identified before, if it is likely to contain part of a gene, and the possible function of any such encoded protein. As the Human Genome Project accelerates towards completion of the genome map, the use of computational tools to achieve these objectives is likely to become increasingly rewarding. The aim of this chapter is to direct the inexperienced scientist toward the most essential programs and to describe their use. More advanced information can be found in recent books which cover the subject comprehensively ***(1,2)***.

The resources for DNA analysis are available on the World Wide Web and are most easily accessed through centers such as the Human Genome Mapping Project Resource Centre in the UK (http://www.hgmp.mrc.ac.uk) or via Workbench at the National Center for Supercomputing Applications (NCSA) in the United States (http://workbench.sdsc.edu). These centers provide an interface to a very wide and comprehensive range of programs covering all aspects of DNA and protein analysis. Registration of users is required before use. On-line help is available and users are directed through complex programs using sensible defaults. Further assis-

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

tance by e-mail is generally provided, if necessary. The same programs can also be used independently, and their locations are provided in the following subheading.

2. BLAST Searches for Sequence Homology

BLAST (Basic Local Alignment Search Tool) is a group of programs that search available databases for similarities with either DNA or protein query sequences *(3–6)*. The BLAST algorithms search for identity in a short region (default = 11 bases), which is extended if possible into regions of close similarity. They are therefore not suitable for testing short sequences such as primers because very close matches may not be identified. The most recent version allows gaps to be included between matched sequences *(7)*. BLAST is effective in detecting similarity between sequences that share only an isolated region in common. The programs are designed for speed (usually taking minutes) with minimal sacrifice of sensitivity, whereas FASTA is more sensitive but takes hours. BLAST is therefore the main search workhorse.

2.1. Performing a BLAST Search

1. Connect to http://www.ncbi.nlm.nih.gov/BLAST/. The interface to the BLAST programs is friendly and intuitive. The Overview, Manual, and Frequently Asked Questions links here provide useful help if necessary.
2. Click the Basic BLAST search link below BLAST 2.1.
3. On the next screen, select your preferred BLAST program from the drop down menu. Information about the choices can be obtained from the Program link. The default, blastn, is used most often.
   - **blastn** compares a nucleotide query sequence against a nucleotide sequence database.
   - **blastx** compares a nucleotide query sequence translated in all six reading frames against a protein sequence database.
   - **tblastx** compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. This program cannot be used with the nr database.

   Other options are available for protein sequences.

4. Select the database to search from the drop down menu. Information about the choices can be obtained from the Database link. The default, nr, is often the preferred option. The following may be of interest.
   - **nr** All Non-redundant GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1, or 2 HTGS sequences).
   - **month** All new or revised GenBank+EMBL+DDBJ+PDB sequences released in the last 30 d; useful to update BLAST searches.
   - **dbest** Nonredundant Database of GenBank+EMBL+DDBJ EST Divisions.
   - **dbsts** Nonredundant Database of GenBank+EMBL+DDBJ STS Divisions.
   - **htgs** Unfinished High-Throughput Genomic Sequences: phases 0, 1, and 2 (finished, phase 3 HTG sequences are in nr).
   - **vector** Vector subset of GenBank(R), NCBI, in ftp://ncbi.nlm.nih.gov/blast/db/
5. The default to filter your query sequence for low compositional complexity regions should be used to reduce spurious matches. Filtered regions will appear as strings of N's in the DNA sequence in the output (or X's in protein sequences).
6. Enter or paste in your input sequence into the box in FASTA format. Information on this format is available by clicking the FASTA link below the input data box. It is not necessary to include the single line description. This format is case insensitive. Sequence can be pasted directly from some other formats without removing digits and spaces. BLAST searches can be performed on Genbank data by entering the accession number and altering the format box from "Sequence in FASTA" to "Accession or GI" in the drop down box.
7. The results of the BLAST search can be received by e-mail by ticking the box and inserting an e-mail address.
8. The search is initiated by clicking either of the search buttons on this page.
9. On the next screen, there is an estimate of the time before the results will be available. The "Format Results" button can be clicked earlier than this without penalty, and the screen will automatically update until the results appear.

The results of the BLAST search appear with a graphical overview that is color coded to indicate the closeness of match, with the

best hits appearing in red. Alignments to the same database sequence are connected by a striped line. The identity of each match is revealed by placing the mouse over the colored hit sequence; clicking on it takes the user down the page to the associated alignment. The diagram of hits is followed by a list of sequences (with short descriptive titles) in the database with similarity to the input sequence. The list is ranked by score (S) and E value (expect), with the closest matches at the top. E is the number of hits expected by chance searching a database of a particular size and decreases exponentially with the score (S) assigned to a match. Next follows information about the percent identity in the region of similarity of each match and a display of the relationship between the query and the related sequence from the database. Only the overlapping regions with partial or complete identity are shown, along with their base numbers in the two sequences.

2.2. Interpretation of the BLAST Search Results

The most interesting data fall into three categories.

1. A >97% identity between a significant part of the input sequence and the database entry. All or part of your DNA fragment has been sequenced before, or there may be a match between genomic DNA and its equivalent cDNA. Differences between the sequences are probably due to sequencing errors, for example, involving strings of a single base.
2. An 80–90% similarity may indicate homology with a gene in a different species, or with pseudogenes. Identification of a mouse homolog of a human gene can often provide valuable information about gene organization and function. Critical regions of genes are usually conserved across species.
3. Low level similarity is found between members of a supergene family. Genes that share functional motifs will also show similarity within a short region. A cluster of weak matches with functionally related genes can give a good indication of possible function, particularly if identified using **blastx**.

Remember that many of the DNA and protein entries in the databases have errors—you must interpret your results with care! A

very large number of sequences are contaminated with fragments of vector, usually at the ends but occasionally internally. It is important to remove vector sequence from your data before searching. Sequences that contain Alu or other repetitive elements will also give rise to a vast number of hits that may prevent more interesting weak matches from being identified. If the majority of hits involve a match with one particular region within a sequence, it may be useful to search again after removal of this part of the sequence. Some researchers prefer to BLAST open reading frame regions only to avoid hits that are unlikely to be relevant (**ORF Finder**: http://www.ncbi.nlm.nih.gov/gorf/gorf.html). This is unwise unless the sequence is all of good quality. The BLAST programs for translated protein (e.g., blastx) can be used to identify possible sequencing errors that cause frameshifts, although this can be achieved more easily using the fserr program of GRAIL via NIX (*see* **Subheading 3.2.**).

2.3. Alternative Options Within BLAST

There are several interesting options for database searching on the main BLAST page (http://www.ncbi.nlm.nih.gov/BLAST/).

1. The **advanced BLAST** program allows the defaults to be altered. This is seldom necessary, but can be useful on rare occasions. For example, when searching with short sequences, the word size default of 11 nt should be reduced. The penalties for introducing and extending gaps can also be amended.
2. The new **BLAST 2 sequences** option allows comparison of any two sequences, which are entered into the boxes either as sequence data or accession numbers. This is useful for identifying the positions of introns if genomic and cDNA sequences are compared, and for finding regions of difference between a gene and related pseudogene, to improve selection of gene-specific primers.
3. **Multiple sequence alignment** can be achieved with programs such as MACAW from NCBI (ftp://ncbi.nlm.nih.gov/pub/macaw/) and EMMA from the EMBOSS group of applications (http://www.sanger.ac.uk/Software/EMBOSS/Apps/) which interfaces with ClustalW.

3. Software for GENE STRUCTURE ANALYSIS

3.1. Exon Identification

Several programs are available to search a DNA sequence for potential exons. The most useful are:

- **MZEF** developed by Michael Zhang *(8)* (http://www.cshl.org/genefinder), click, for example, the human or mouse link for the data input page; and the group of Genefinder programs *(9)*, including HEXON, FEX, and FGENEH, from Baylor College of Medicine Computational Biology Group (http://mbcr.bcm.tmc.edu/Guide/Genefinder/form.html for input data form).
- **GRAIL 2** from the Oak Ridge National Laboratory *(10)* (http://compbio.ornl.gov/Grail-1.3/ for input data form).

These software-trapping programs use a variety of parameters to predict the positions of exons, such as exon length of open reading frame, 5' and 3' splice site sequences and occasionally branch-site information. Many can detect only internal exons and require more than 50 bp of flanking intronic sequence, although FEX attempts to find 5' and 3' exons. Short exons (<100 bp) are at high risk of being ignored. GRAIL also searches for poly(A) sites, CpG islands, and frameshift errors. The programs use complex algorithms and scoring systems, but are generally straightforward to use. They have a range of accuracy of 50–80% and are best limited to analysis of sequences of <150 kb. All are capable of identifying false exons and missing existing ones, so it is advisable to analyze data by several programs. Exons that are supported strongly by more than one program are most worthy of follow-up.

3.2. Simultaneous Gene Structure Analysis Using NIX

NIX, which is provided by the UK HGMP (follow NIX link under Bioinformatics on http://www.hgmp.mrc.ac.uk/), is a most impressive tool to achieve simultaneous analysis of several programs including MZEF, FEX, HEXON, GRAIL, GENEMARK, GENEFINDER, FGENE, BLAST (against many databases), Polyah, RepeatMasker, and tRNAscan on a DNA sequence of interest. Repetitive elements in the

input sequence are automatically masked using REPEATMASKER prior to analysis. The optimum length of DNA sequence for analysis is 20–50 kb, and longer sequences should be cut into overlapping segments within this size range. The lower limit is 100 bp. The NIX interface is extremely intuitive. The introduction to NIX (http://www.hgmp.mrc.ac.uk/NIX/) provides invaluable information about the operation of NIX and interpretation of its results. The user has no control of the intelligent defaults used for DNA analysis, but has minimal need to understand the complexities of each individual program.

3.2.1. Running NIX

1. Register with UK HGMP. (Follow Registration link on http://www.hgmp.mrc.ac.uk/)
2. Click NIX link on the menu at http://www.hgmp.mrc.ac.uk/, and Run NIX now! on the following screen.
3. Enter the DNA sequence either by file (100–150 kb) or by pasting (<20 kb), and provide a short description in the boxes provided. Check if genomic or mRNA/cDNA/EST and the origin of the DNA. Provide an email address for notification when the results are ready. Click "Start the Analysis" button and go home! Results are usually not available until the following day.
4. Results can be viewed by following the View your NIX Results now! link on the NIX page (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/). An excellent explanation of how to interpret the results is provided here on the Information on the results viewer link. Viewing the results of several exon-finding programs side by side allows consensus and near-consensus exons to be identified. BLAST search results are also shown. In all cases, there is color-coding of information, with brighter hues indicating more confident predictions, for example, of exons or BLAST matches. It is useful to reformat the data after removing some of the less interesting results, to assist with viewing of at least one dialogue box that indicates information on the feature under the mouse.

3.3. Other DNA Analysis Programs

Many computer programs have been developed to perform a wide range of simple but often time-consuming procedures on DNA

sequences. These include reverse complementing sequence, identifying motifs, open reading frames, and primer design. They are freely available by following informatics links on Web pages of most large genome centres. Some interesting URLs are listed below.

The Sanger Center	http://www.sanger.ac.uk/
Stanford Human Genome Ctr.	http://www-shgc.stanford.edu/
European Molecular Biology Laboratory	http://www.embl-heidelberg.de/
Whitehead Institute	http://www-genome.wi.mit.edu/
Baylor College of Medicine	http://www.hgsc.bcm.tmc.edu/SearchLauncher/
U.S. Natl. Ctr. for Biotechnol. Info.	http://www.ncbi.nlm.nih.gov/

References

1. Baxevanis, A. and Ouellette, B. F. F. (eds.) (1998) *Bioinformatics. A Practical Guide to the Analysis of Genes and Proteins*, Wiley-Interscience.
2. Bishop, M. J. (ed.) (1998) *Guide to Human Genome Computing*, Academic Press.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410.
4. Karlin, S. and Altschul, S. F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* **87,** 2264–2268.
5. Karlin, S. and Altschul, S. F. (1993) Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci. USA* **90,** 5873–5877.
6. Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994) Issues in searching molecular sequence databases. *Nat. Genet.* **6,** 119–129.
7. Altschul, S. F., Madden, T. L., Schaffer, A. A., et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.
8. Zhang, M. Q. (1997) Identification of protein coding regions in the human genome by quadratic discriminant analysis [published erratum appears in *Proc. Natl. Acad. Sci. USA* (1997) **94(10),** 5495]. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 565–568.

9. Solovyev, V. V., Salamov, A. A., and Lawrence, C. B. (1994) Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* **22,** 5156–5163.
10. Uberbacher, E. C. and Mural, R. J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88,** 11,261–11,265.

15

DNA Sequencing by Capillary Array Electrophoresis

Norman J. Dovichi and JianZhong Zhang

1. Introduction

On May 9, 1998, Applied Biosystems announced that they would introduce a new instrument based on "breakthrough DNA analysis technology" ***(1)***. Over 700 orders for the new sequencer were received in the first five months of production, which make this instrument one of the most successful new products in the history of analytical instrumentation. This new instrument is based on capillary array electrophoresis (CAE) with a sheath-flow cuvet detector. This detection technology provides nearly 100% duty cycle, where every capillary is monitored simultaneously without the need for a scanning detector system. By use of this instrument, the *Drosophila melanogaster* sequence will be completed in less than 5 mo ***(2)***, and the sequence of the human genome will be substantially complete by the time that this book is in print.

DNA sequencing is used to identify genes and mutations, to confirm successful site directed mutagenesis, and to analyze phage-display libraries. The vast majority of DNA sequence is determined by use of the chain-termination method developed by Sanger's group in 1977 ***(3)***. In the original method, a DNA sequencing ladder is

From: *Methods in Molecular Biology, vol. 167: DNA Sequencing Protocols, 2nd ed.*
Edited by: C. A. Graham and A. J. M. Hill

synthesized, radioactively labeled, separated by polyacrylamide gel electrophoresis, detected by autoradiography, and interpreted by a skilled technician. This technology, although cumbersome and manually intensive, has remained the method of choice for DNA sequencing. Several advances have been made to improve the technology by replacing the use of radioactive labels, autoradiographic detection, and manual data interpretation with fluorescent labels, laser-induced fluorescence detection, and computer-based automatic data analysis *(4–6)*. Several instruments based on these principles have been commercialized.

1.1. Automated Sequencing

The commercialization of automated DNA sequencers is an important event in biotechnology. A group of visionary biologists realized that this new technology would allow the completion of very large DNA sequencing projects, and the Human Genome Project was initiated in 1988, two years after the first report of an automated DNA sequencer *(7)*. The National Institutes of Health of the United States has provided the lion's share of public funding for the Human Genome Project, which has grown into an international effort of global proportions. The Welcome Foundation in Great Britain has been notable in their support of the Human Genome Project. In the last year, Celera has become an important private-sector company active in very large-scale sequencing efforts.

The original goals of the Human Genome Project envisioned that the entire genome would be sequenced by the year 2005. Improved DNA sequencing technology has become available that dramatically speeds the sequencing effort. It is estimated that 90% of the human genome will be sequenced in the year 2000 *(8)*. The advances in DNA sequencing technology include the development of large-scale shot-gun sequencing *(9)*, highly automated DNA sequencers, and powerful data processing algorithms. This chapter focuses on the latest generation of highly automated DNA sequencer, which is based on CAE.

1.2. Sequencing in Slab Gels

Conventional DNA separations are performed using slab gel electrophoresis that are coated with a ~200-μm thick layer of crosslinked polyacrylamide. The polymer is allowed to polymerize between two glass plates. The relatively thick, crosslinked polymer is not ideal for DNA separations. From a practical point of view, the thick, rigid polymer must be scraped from the glass plates after every separation. Once the tedious scraping step is completed, the plates must be washed and reassembled. The acrylamide mixture is then poured between the plates and allowed to polymerize. DNA is manually loaded into discrete wells at the top of the gel and the plates are inserted into the sequencing instrument. This procedure is very expensive in large-scale sequencing centers, where the personnel cost associated with handling the sequencing plates dominates the cost of DNA sequencing.

On a more fundamental point of view, the thick gels used in classic sequencing are not able to rapidly generate DNA sequence. DNA sequencing speed is related to the electric field applied to separate the fragments. To first approximation, the sequencing speed increases linearly with the applied electric field. However, application of a voltage across a material with high conductivity results in the generation of heat; this effect is called Joule heating. Because of their thickness, classic slab gels can not efficiently radiate heat, and Joule heating limits the maximum applied electric field to rather modest levels, which results in rather slow sequencing speeds.

2. Capillary Electrophoresis (CE)

2.1. Sequencing by CE

CE has replaced conventional electrophoresis and has dramatically increased the speed of DNA sequencing. These capillaries are typically 50 μm in inner diameter, are about 30 cm in length, and are made from high-purity fused silica. The small inner diameter of the fused silica capillaries result in excellent thermal properties, which reduces Joule heating to negligible levels and allows the use of

extremely high electric fields for very rapid separations of DNA sequencing fragments. However, the phenomenon of biased reptation with orientation is observed at high electric field, and this degrades the separation of longer sequencing fragments ***(10)***. In practice, very high electric fields are not useful in DNA sequencing applications. Instead, the most important property of the fused silica capillaries is that they are highly flexible and easily incorporated into an automated instrument. It is the ease of automation, rather than fundamental thermal properties, that have resulted in the successful development of advanced sequencing instruments.

2.2. Capillary Array Electrophoresis

Whereas the performance of CE is superior to that of conventional slab-gel electrophoresis, a single-capillary instrument does not offer significant advantages compared with a multilane slab-gel system. Instead, it is necessary to operate an array of capillaries to obtain throughput that is comparable with conventional electrophoresis systems.

The first report of CAE for DNA sequencing appeared in 1990 ***(11)***. The detector of that instrument scanned across the capillaries, recording the fluorescence signal sequentially from each capillary. Other groups have also used this approach of scanning a detector across the capillary array for DNA detection. However, the scanning of capillaries for DNA detection suffers from a limited duty cycle period: although one capillary is being interrogated by the detector, DNA is also migrating from the other unscanned capillaries, so that most DNA is not being detected. Limited duty cycle can be a severe restraint, for example, each capillary in a 96-capillary array is probed for 1% of the time, and 99% of the DNA is lost without detection.

2.3. Continuous Capillary Monitoring

The more successful capillary array sequencers rely on continuous monitoring of each capillary. Two instruments are of particular interest, one from the laboratory of H. Kambara at Hitachi and the other from this group ***(12,13)***. Both instruments rely on a sheath-

flow cuvet to simultaneously monitor fluorescence from a linear array of capillaries. This technology allows one laser beam to simultaneously illuminate samples migrating from all capillaries. An optical system images the fluorescence from each capillary onto a charged coupled device (CCD) camera or an array of photodiodes. In this way, fluorescence from each capillary is monitored continuously, and the instrument's duty-cycle approaches 100%.

2.3.1. Single Capillary Sheath-Flow Detection

The sheath-flow detector was originally developed for use in flow cytometry, where a dilute suspension of fluorescently labeled cells is pumped into the center of a flowing sheath stream under laminar flow conditions ***(14)***. The cellular suspension flows through the center of the flow chamber and intersects a focused laser beam. If the sheath buffer is similar to the buffer in which the cells are suspended, then there is little refractive index difference between the two streams, and no light is scattered at their interface. By use of a flow chamber with flat windows, the fluorescence signal from each cell may be measured with very low background signal from scattered laser light. Flow cytometry operates at high flow rates in order to process large numbers of cells, so that the cellular population may be characterized with high precision.

Dick Keller's laboratory at Los Alamos Scientific Laboratory pioneered the use of the sheath-flow cuvet as a high sensitivity detector for fluorescent dyes ***(15,16)***. In favorable cases, single fluorescent molecules have been detected as they passed through the focused laser beam ***(17)***. My laboratory has pioneered the use of the cuvet as a detector in CE ***(18–20)***. The outstanding detection performance of the cuvet, coupled with the very small probe volume produced at the intersection of the laser beam and sample stream, make the detector well suited for application in CE. However, the volumetric flow rate employed in CE is many orders of magnitude lower than that employed in flow cytometry, and care is required to optimize the performance of the detector.

In these instruments, the separation capillary is inserted into a square flow chamber, illustrated in **Fig. 1**. We typically use a

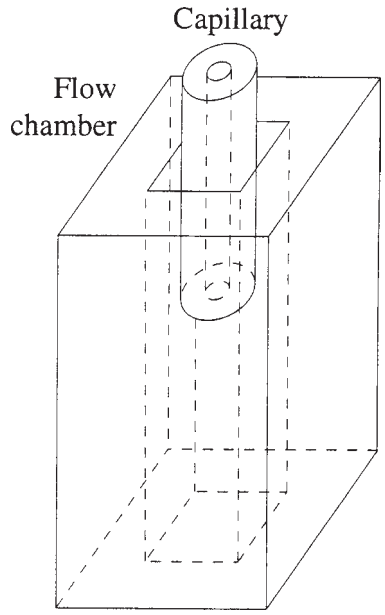


Fig. 1. Single capillary sheath-flow cuvet. A fused silica capillary is placed inside of a square quartz flow chamber.

150-μm od fused silica capillary for electrophoresis. The flow chamber is typically 165-μm square and 1.5 cm long. The windows on the cuvet are usually 1 mm thick. Sheath buffer is pumped at a low flow rate from the top of the cuvet to surround the sample stream as it migrates from the capillary. A low-power laser beam is focused into the cuvet, typically forming a 50-μm spot about 150-μm downstream from the capillary exit (*see* **Fig. 2**). Since fluorescence is detected in the flow chamber and not in the capillary, this detector is a rare example of a post column detector in CE. Fluorescence is collected with a high-efficiency microscope objective, spectrally filtered to reduce scattered laser light, imaged onto a mask to further block scattered laser light, and detected with a high-efficiency photomultiplier tube. The flow chamber is held in a stainless-steel fixture, which is held at ground potential and which completes the electrophoresis circuit.

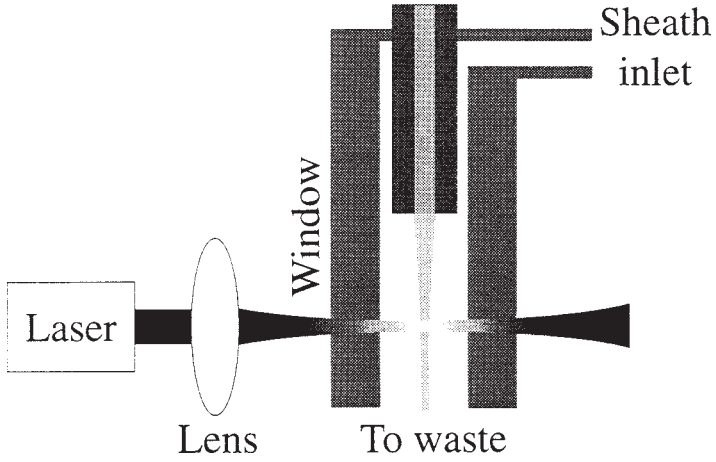


Fig. 2. Single capillary sheath-flow cuvet instrument. Sheath fluid draws the analyte into a thin stream in the center of the flow chamber. A laser beam is focused beneath the capillary tip on the sample stream. A lens collects the fluorescence emission signal, which is then spectrally filtered, then imaged onto an aperture and detected with a photomultiplier tube.

2.3.2. Sensitivity of Sheath-Flow Cuvet Detection

The sheath-flow cuvet has been used as a detector for the CE analysis of zeptomole amounts of fluorescently labeled amino acids, peptides, proteins, monosaccharides, oligosaccharides, and oligonucleotides ***(21–26)***. The detection limits are routinely in the yoctomole range. We have also detected single molecules of β-phycoerythrin after CE separation with this detector ***(27)***.

In 1990 this group ***(28)*** reported the first application of CE with sheath-flow detection for the separation of DNA sequencing fragments, and the first four-color sequencing application was reported in the following year ***(29)***. The sheath-flow cuvet has proven to be quite reliable and able to generate long sequencing read-lengths when combined with high-temperature electrophoresis separation of sequencing fragments ***(30,31)***.

A reviewer of one of our proposals naively felt that the sheath-flow cuvet could not function as a detector in DNA sequencing

applications. Because there is no electroosmotic flow from the gel-filled capillary to carry analyte into the cuvet, the reviewer assumed that the analyte would remain trapped at the capillary outlet and would not migrate into the flowing sheath stream. Of course, diffusion rapidly carries analyte into the sheath stream, where it is efficiently swept downstream to the fluorescence excitation volume. Fluorescence must be resolved into spectral bands for DNA sequencing applications. We have reported the use of both a rotating filter wheel and a dichroic filter assembly for this purpose.

2.3.3. Capillary Array Sheath-Flow Detection

As part of his PhD thesis, JianZhong Zhang developed a capillary array DNA sequencer based on a sheath-flow cuvet ***(32)***. Simultaneously and independently, H. Kambara's group developed a similar instrument ***(33,34)***. In these instruments, a linear array of capillaries is inserted into a rectangular sheath-flow cuvet, illustrated in **Fig. 3**. Sheath fluid is pumped through the interstitial space between the capillaries and entrains the DNA sequencing fragments as discrete streams, with one stream per capillary. A laser beam is focused into the cuvet and skims beneath the capillary tips, as shown in **Fig. 4**, exciting fluorescence from all of the capillaries, simultaneously. When observed from the front, the detector generates a set of fluorescent spots, with one spot beneath each capillary and with the spots separated by the outer diameter of the capillary. **Reference *35*** presents a color photograph of a cuvet in operation. A high numerical aperture lens is used to collect the fluorescence and to image each fluorescent spot onto a discrete photo-detector. In our original systems, we used a set of fiber-optic coupled avalanche photodiodes, which were operated in the photon counting mode. More recently, we have used a CCD camera to image the fluorescence. The laser beam traverses the set of sample streams that are migrating from each capillary. Fortunately, there is a negligible loss of laser power in traversing the cuvet. The analyte concentration is quite low and the optical path-length across each sample

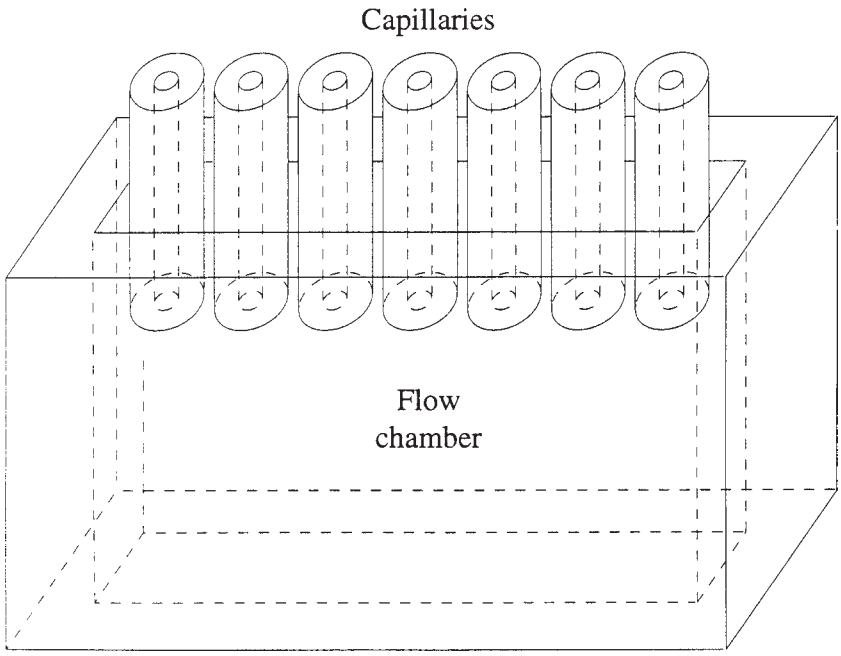


Fig. 3. A linear-array capillary, sheath-flow cuvet. A linear array of fused silica capillaries is placed inside of a rectangular glass flow chamber.

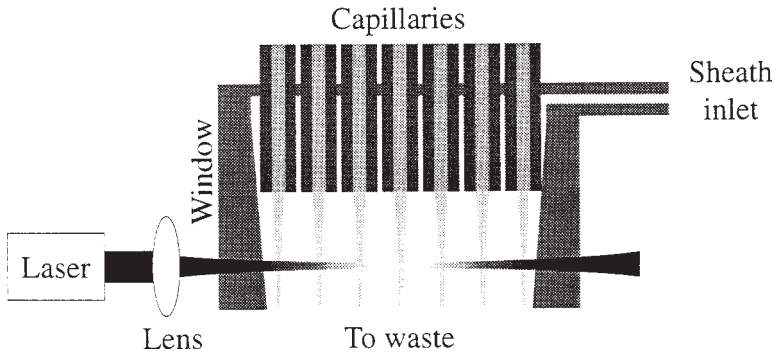


Fig. 4. Capillary linear-array sheath-flow cuvet instrument. Sheath fluid draws the analyte into thin streams in the center of the flow chamber, with a single stream produced downstream from each capillary. A laser beam is focused beneath the capillary tips on the sample streams. A lens collects the fluorescence emission signal, which is then spectrally filtered and detected with either an array of photodiodes, or with a CCD camera.

stream is small. As a result, the loss of laser power owing to absorbance of the beam by the sample is much less than 1%, even for the highest concentration primer peak that migrates early in the separation.

2.4. Capillary Spacing

If the capillaries are evenly spaced within the cuvet, then the sample streams will also be evenly spaced. If the capillaries are not evenly spaced, the hydrodynamic flow of the sheath fluid between the capillaries will not be uniform. Instead, the sheath-flow rate will be higher in the region where the capillaries are spaced wider than normal and the sheath-flow rate will be lower in the regions where the capillaries are closer together. These differences in the rate of sheath flow, cause the sample streams to be drawn together where the capillaries are close together. When illuminated by the laser beam, the fluorescent spots will not be uniformly spaced, and alignment of the detectors with the images of the spots will be difficult and irreproducible. We find that the capillary spacing needs to be within ~5 μm to ensure even spacing of the fluorescence spots.

There are two methods of generating evenly spaced capillaries. In the first case, the cuvet is designed so that the capillaries are in intimate contact. As a result, their center-to-center spacing is determined by the outer diameter of the capillary, which is typically held to within 1 μm across a 100-m long reel of capillary. However, the reel-to-reel dimensions of the capillary can vary by more than 10 μm. To avoid the custom manufacture of a cuvet for each reel of capillary, we developed a wedge-shaped cuvet, wherein the walls of the cuvet are tapered. The capillaries are forced into contact as they are inserted into the tapered cuvet.

In the second approach to ensuring uniform capillary spacing, a set of fingers can be micromachined in the inner wall of the cuvet. These features are raised with respect to the cuvet walls and act to hold the capillaries on uniform centers.

3. Commercial CAE Instruments

The CAE instrument has been commercialized by Applied Biosystems as the model 3700 DNA sequencer ***(1,34,36)***. The instrument operates with 96 capillaries in its sheath-flow detector for convenient mating with a 96-well microtiter plate. There are also eight spare capillaries in the instrument. A robot automatically and rapidly transfers samples from a 96-well microtiter plate to an injection block, where samples are loaded onto the capillaries. Sample loading, sieving matrix replacement, electrophoresis, data collection, and base-calling are all automated, so that the instrument can run 24 h without intervention.

The commercial instrument uses a concave spectrograph to image fluorescence onto a CCD camera, which records the fluorescence spectrum generated by each capillary. This spectral data is monitored continuously, which ensures nearly a 100% duty cycle for detection. The data is then processed to obtain sequence data. Unlike slab-gel data, there is no need for lane-finding algorithms to process the data; the capillary format ensures that each sample is imaged to a specified location on the camera.

Although the instrument is specified as producing 550 bases of sequence data at 98.5% accuracy, the performance of this instrument has exceeded expectation ***(2)***. Celera has announced that they have generated over 500 million bases of DNA sequence from the *Drosophila* genome in 1 million electropherograms. It is expected that this performance will improve as the sequencing software is optimized.

4. Conclusions

The human genome project is the most ambitious and important effort in the history of biology. It will provide the complete genetic blueprint for human life, and will provide important insights into human health and development. High-throughput DNA sequencers based on CAE with a sheath-flow detector will do the lion's share of the sequencing effort. The highly flexible, automated, and effi-

cient CE technology with the high sensitivity and high duty-cycle sheath-flow cuvet is making the human genome project a reality. This technology is also one of the best selling products in the history of analytical instrumentation, with sales of over 700 instruments in the first few months of production. Combined, the sales of these instruments have a value of $200 million (US) ***(37)***.

There will be many more applications of this high-throughput technology beyond the initial sequencing of the human genome. Comparative genomics will require that the DNA sequence from many individuals be obtained and analyzed. Human disease diagnosis and prognosis will become very important in the near future, particularly as genetic markers are found that are associated with cancer stage and prognosis; these markers will be routinely sequenced in clinical laboratories. The genome from other vertebrate organisms will be sequenced to gain insight into evolution. The genomes of agriculturally important plants and animals will be sequenced to provide better control and manipulation of economically important traits. Plant and animal breeding programs will require that large numbers of organisms be genotyped.

We are at the threshold of a new era in the biological sciences. This era is based on genomic information that was unimaginable two decades ago. The development of novel analytical technology has made the genome project possible and is decreasing dramatically the cost of the sequencing effort. This is a very exciting time in the biological sciences!

Acknowledgments

This work was supported by the Canadian Genetic Diseases Network, the Natural Sciences and Engineering Research Council, and SCIEX.

References

1. http://www.pecorporation.com/press/prc5448.html
2. http://www.pecorporation.com/press/prccorp0728.html

3. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci. USA* **74,** 5463–5467.
4. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679.
5. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238,** 336–341.
6. Ansorge, W., Sproat, B. S., Stegemann, J., and Schwager, C. (1986) A non-radioactive automated method for DNA sequence determination. *J. Biochem. Biophys. Meth.* **13,** 315-317.
7. http://www.nhgri.nih.gov/HGP
8. http://www.nhgri.nih.gov:80/NEWS/Finish_sequencing_early/Intent_to_finish_sequencing_ early.html
9. Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998) Shotgun sequencing of the human genome. *Science* **280,** 1540–1542.
10. Viovy J.-L. and Duke, T. (1993) DNA electrophoresis in polymer solutions: Ogston sieving, reptation and constraint release. *Electrophoresis* **14,** 322–329.
11. Zagursky, R. J. and McCormick, R. M. (1990) DNA sequencing separations in capillary gels on a modified commercial DNA sequencing instrument. *BioTechniques* **9,** 74–79.
12. Dovichi, N. J. and Zhang, J. Z. (1995) Multiple capillary biochemical analyzer. US Patent 5,439,578; 8 August 1995.
13. Takahashi, S. and Kambara, H. (1996) DNA detector and DNA detection method. US Patent 5,529,679; June 25, 1996.
14. Steinkamp, J. A. (1984) Flow cytometry. *Rev. Sci. Instrum.* **55,** 1375.
15. Dovichi, N. J., Martin, J. C., Jett, J. H., Trkula, M., and Keller R. A. (1984) Laser-induced fluorescence of flowing samples as an approach to single-molecule detection in liquids. *Anal. Chem.* **56,** 348–354.
16. Dovichi, N. J., Martin, J. C., Jett, J. H., and Keller R. A. (1983) Attogram detection limit for aqueous dye samples by laser-induced fluorescence. *Science* **219,** 845–847.
17. Hguyen, D. C., Keller, R. A., Jett, J. A., and Martin, J. C. (1987) Detection of single molecules of phycoerythrin in hydrodynamically

focused flows by laser-induced fluorescence. *Anal. Chem.* **59,** 2158–2161.

18. Cheng, Y. F. and Dovichi, N. J. (1988) Subattomole amino acid analysis by capillary zone electrophoresis and laser-induced fluorescence. *Science* **242,** 562–564.
19. Cheng, Y. F., Wu, S., Chen, D. Y., and Dovichi, N. J. (1990) Interaction of capillary zone electrophoresis with a sheath flow cuvette detector. *Anal. Chem.* **62,** 496–503.
20. Wu, S. and Dovichi, N. J. (1989) High-sensitivity fluorescence detector for fluorescein isothiocyanate derivatives of amino acids separated by capillary zone electrophoresis. *J. Chromatogr.* **480,** 141–155.
21. Zhao, J. Y., Chen, D. Y., and Dovichi, N. J. (1992) Low-cost laser-induced fluorescence detector for micellar capillary zone electrophoresis: detection at the zeptomol level of tetramethylrhodamine thiocarbamyl amino acid derivatives. *J. Chromatogr.* **608,** 117–120.
22. Zhao, J. Y., Waldron, K. C., Miller, J., Zhang, J. Z., Harke, H. R., and Dovichi, N. J. (1992) Attachment of a single fluorescent label to peptides for determination by capillary zone electrophoresis. *J. Chromatogr.* **608,** 239–242.
23. Pinto, D. M., Arriaga, E. A., Craig, D., Angelova, J., Sharma, N., Ahmadzadeh, H., Dovichi, N. J., and Boulet, C. A. (1997) Picomolar assay of native proteins by capillary electrophoresis - precolumn labeling, sub-micellar separation, and laser induced fluorescence detection. *Anal. Chem.* **69,** 3015–3021.
24. Zhang, Y., Arriaga, E., Diedrich, P., Hindsgaul, O., and Dovichi, N. J. (1995) Nanomolar determination of aminated sugars by capillary electrophoresis. *J. Chromatogr.* **716,** 221–229.
25. Le, X.-C., Scaman, C., Zhang, Y., Zhang, J. Z., Dovichi, N. J., Hindsgaul, O., and Palcic, M. M. (1995) Analysis by capillary electrophoresis-laser-induced fluorescence detection of oligosaccharides produced from enzyme reactions. *J. Chromatogr.* **716,** 215–220.
26. Figeys, D., Arriaga, E., Renborg, A., and Dovichi, N. J. (1994) Use of the fluorescent intercalating dyes POPO-3, YOYO-3 and YOYO-1 for ultrasensitive detection of double stranded DNA separated by capillary electrophoresis with hydroxypropylmethyl cellulose and non-cross-linked polyacrylamide. *J. Chromatogr.* **669,** 205–216.
27. Chen, D. Y. and Dovichi, N. J. (1996) Single-molecule detection in capillary electrophoresis: molecular shot noise as a fundamental limit to chemical analysis. *Anal. Chem.* **68,** 690–696.

28. Swerdlow, H., Wu, S., Harke, H., and Dovichi, N. J. (1990) Capillary gel electrophoresis for DNA sequencing—laser-induced fluorescence detection with the sheath flow cuvette. *J. Chromatogr.* **516,** 61–67.
29. Swerdlow, H., Zhang, J. Z., Chen, D. Y., Harke, H. R., Grey, R., Wu, S., Fuller, C., and Dovichi, N. J. (1991) Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. *Anal. Chem.* **63,** 2835–2841.
30. Zhang, J. Z., Fang, Y., Hou, J. Y., Ren, H. J., Jiang, R., Roos, P., and Dovichi, N. J. (1995) Use of non-cross-linked polyacrylamide for four-color DNA sequencing by capillary electrophoresis separation of fragments up to 640 bases in length in two hours. *Anal. Chem.* **67,** 4589–4593.
31. Yan, J. Y., Best, N., Zhang, J. Z., Ren, H. J., Jiang, R., Hou, J., and Dovichi, N. J. (1996) The limiting mobility of DNA sequencing fragments for both cross-linked and noncross-linked polymers in capillary electrophoresis: DNA sequencing at 1200 V cm^{-1}. *Electrophoresis* **17,** 1037–1045.
32. Zhang, J. Z. (1994) DNA sequencing by single and multiple capillary gel electrophoresis. *PhD thesis*, Department of Chemistry University of Alberta.
33. Kambara, H. and Takahashi, S. (1993) Multiple-sheath flow capillary array DNA analyzer. *Nature* **361**, 565-566.
34. Kamahori, M. and Kambara, H. (2000) Capillary array electrophoresis analyser, in *Capillary Electrophoresis of Nucleic Acids,* Vol. 2 (Mitchelson, K. R. and Cheng, J., eds.), Humana Press, Totowa, NJ, pp. 271–287.
35. Anonymous comment. *Science* 1998, **280,** 995.
36. http://www2.perkin-elmer.com/ga/3700/features.html
37. http://www.pecorporation.com/press/dlw042999.html