0273-2289/93/3801-02--0147$03.60

DNA Sequencing

Recent Innovations and Future Trends

H. G. GRIFFIN* AND A. M. GRIFFIN

AFRC Institute of Food Research, Norwich Laboratory, Norwich Research Park, Colney, Norwich, NR4 7UA England

ABSTRACT

Determination of the sequence of DNA is one of the most important aspects of modern molecular biology. New sequencing methods currently being developed enable DNA sequence to be determined increasingly faster and more efficiently. One of the major advances in sequencing technology is the development of automated DNA sequencers. These utilize fluorescent rather than radioactive labels. A laser beam excites the fluorescent dyes, the emitted fluorescence is collected by detectors, and the information analyzed by computer. Robotic work stations are being developed to perform template preparation and purification, and the sequencing reactions themselves. Research is currently in progress to develop the technology of mass spectrometry for DNA sequencing. Success in this endeavor would mean that the gel electrophoresis step in DNA sequencing could be eliminated.

A major innovation has been the application of polymerase chain reaction (PCR) technology to DNA sequence determination, which has led to the development of linear amplification sequencing (cycle sequencing). This very powerful yet technically simple method of sequencing has many advantages over conventional techniques, and may be used in manual or automated methods. Other recent innovations proposed recently to increase speed and efficiency include multiplex sequencing. This consists of pooling a number of samples and processing them as pools. After electrophoresis, the DNA is transferred to a membrane, and sequence images of the individual samples are obtained by sequential hybridizations with specific labeled oligonucleotides. Multiplex DNA sequencing has been used in conjunction with direct blotting electrophoresis to facilitate transfer of the DNA to

*Author to whom all correspondence and reprint requests should be addressed.

a membrane. Chemiluminescent detection can also be used in conjunction with multiplex DNA sequencing to visualize the image on the membrane.

Index Entries: DNA sequencing; sequencing; cycle sequencing; linear amplification sequencing; automated sequencing; fluorescent sequencing; multiplex sequencing; sequence analysis; recombinant DNA technology; genetic engineering; molecular biology.

INTRODUCTION

Determination of the sequence of DNA is one of the most important aspects of modern molecular biology. Since the development in the late 1970s of methods to determine the sequence of DNA (*1,2*), the growth in this area of science has been astronomical. Newer and faster methods have been devised and developed, and a vast amount of sequence data is now available. Many viral genomes have been completely sequenced, including the genome of cytomegalovirus, which consists of 200,000 bp (*3*). Over 75% of the sequence of the bacterium *Escherichia coli* is now available, and more than a quarter of the yeast genome (*Saccharomyces cerevisiae*) has been sequenced (*3*). Projects are also under way to determine the DNA sequence of the fruitfly (*Drosophila melanogaster*), the mouse (*Mus musculus*), a worm (*Caenorhabditis elegans*), a plant (*Arabidopsis thaliana*), and the human (*Homo sapiens*) genome. The entire sequence of the yeast chromosome III was recently deposited with the EMBL data bank in Germany (*4*). This represented 16 months of work by a collaboration of many individual European laboratories. The other yeast chromosomes are currently being sequenced by laboratories in Europe, the United States, Canada, and Japan. The first results of the *C. elegans* project have also recently been published (*5*). The sequence of 121,298 bp from chromosome III of *C. elegans* was determined. This sequence contained many previously unknown genes, and in addition provided new and biologically relevant information on gene density, gene number, DNA duplications, and repeated sequences, as well as the sequence of biologically important genes of known function. The researchers expect to scale up the project in the future and predict the determination of the sequence of 800,000 bp next year and 2,000,000 bp the following year. The realization that the human genome can be sequenced in its entirety has generated great interest and much discussion. Projects to construct physical and genetic maps (a prerequisite to full-scale sequencing) are already under way, and it is expected that the sequencing of the genomes of less complex organisms, such as *E. coli*, yeast, and *C. elegans*, will serve as pilot projects for the human genome project. There is an estimated 3 billion base pairs in the human genome. DNA sequence data from this and other

sequencing projects are stored in international data banks, such as EMBL, GenBank, and DDBJ, so they are readily available to the world's scientific community.

Immensely faster sequencing methods currently being developed are likely to lead to a virtual revolution in molecular biology. The wealth of information obtained from sequence data may well change the entire process of research and discovery in the laboratory. Convention suggests that biological knowledge should come from biochemical and genetic studies of an organism rather than from the analysis of DNA sequence data. Indeed, many scientists argue that the wholesale sequencing of the entire genome of an organism is not only prohibitively costly, but leads only to the discovery of a large number of genes for which functions have not yet been found. Such discoveries are seen by some as being of little consequence, whereas DNA sequencing itself is dismissed as a repetitive and unintellectual pursuit. However, a growing number of researchers are beginning to realize that DNA sequence data, instead of being the end point of a research project, should be the beginning. By a reversal of molecular genetics philosophy, knowledge of the sequence of a gene of unknown function can permit the design of probes and antibodies, which can be used to identify the gene product, its cellular location, and its time of appearance in the developing cell (*6*). A function can be assigned by mutant analysis or by comparison of the deduced amino acid sequence with proteins of known function. Thus, DNA sequencing can act as a catalyst stimulating future research into many diverse areas of science. In the field of medical genetics, for example, instead of using a physiological response to find the gene responsible for a disease and from that to a possible cure, use can be made of the already available sequence of that gene. Many scientists now accept that sequence analysis will provide an increasingly useful approach to the characterization of biological systems, and perceive DNA sequencing as a valuable and often indispensable aspect of their work. Recent technological advances, especially in the area of automated sequencing, have removed much of the drudgery that was once associated with the technique, and modern innovative computer software has greatly simplified the analysis and manipulation of sequence data.

The two original methods of DNA sequencing that were described in 1977 (*1,2*) differ considerably in principle. The enzymatic (or dideoxy chain termination) method of Sanger et al. (*1*) involves the synthesis of a DNA strand from a single-stranded template by a DNA polymerase. The Maxam and Gilbert (or chemical degradation) method (*2*) involves chemical degradation of the original DNA. Both methods produce populations of radioactively labeled polynucleotides that begin from a fixed point and terminate at points dependent on the location of a particular base in the original DNA strand. The polynucleotides are separated by polyacrylamide gel electrophoresis, and the order of nucleotides in the original DNA can be read directly from an autoradiograph of the gel (*7*).

Although both these techniques are still used today, there have been many changes and improvements to the original methods. Although the chemical degradation method is still in use, the enzymatic chain termination method is by far the most popular and widely used technique for sequence determination. This process has been automated by utilizing fluorescent labeling instead of radioactive labeling, and the concepts of polymerase chain reaction (PCR) technology have been harnessed to enable the sequencing reaction to be "cycled" (*8*). Other recent innovations include multiplexing, sequencing by chemiluminescence rather than radioactivity, solid-phase sequencing, and the use of robotic work stations to automate sample preparation and sequencing reactions (*9–11*).

SANGER METHOD

The Sanger method (also called the chain termination, dideoxy, or enzymatic method) involves the synthesis of a new DNA strand by a DNA polymerase enzyme using a single-stranded DNA molecule as a template (*1*). The newly synthesized strand is thus the complement of the template strand. Double-stranded molecules can be used as templates for DNA sequencing, but they must first be denatured to single-stranded form by alkali treatment. The point at which the synthesis of the new strand is initiated is determined by the hybridization of an oligonucleotide to the template. The Sanger method of sequencing depends on the fact that dideoxynucleotides (ddNTPs) are incorporated into the growing strand in the same way as the conventional deoxynucleotides (dNTPs). However, ddNTPs differ from dNTPs in that they lack the 3′-OH group necessary for chain elongation. When a ddNTP is incorporated into the new strand, the absence of the hydroxyl group prevents formation of a phosphodiester bond with the succeeding dNTP, and chain elongation terminates at that position. By using the correct ratio of the four conventional dNTPs and one of the four ddNTPs in a reaction with DNA polymerase, arbitrary incorporation of the ddNTP is achieved, and a population of polynucleotide chains of varying lengths is produced. Synthesis is initiated at the position where the oligonucleotide primer anneals to the template, and each chain is terminated at a specific base (either A,C,G, or T depending on which ddNTP was used). By using the four different ddNTPs in four separate reactions, four populations of polynucleotides are obtained each consisting of a mixture of molecules the lengths of which are determined by the base composition of the original DNA fragment. The complete sequence information can be obtained by electrophoresing the reaction mixtures in a polyacrylamide gel. In conventional manual sequencing, one of the dNTPs is radioactively labeled, so that the information gained

by electrophoresing the four reactions in adjacent tracks of a polyacrylamide gel can be visualized on an autoradiograph.

The original method used the Klenow fragment of DNA polymerase I to synthesize the new strands in the sequencing reactions, and this enzyme is still used today. Other enzymes, such as Sequenase, T7 polymerase, and *Taq* polymerase, are also widely used (*12–14*). 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.

MAXAM AND GILBERT METHOD

The Maxam and Gilbert (or chemical) procedure for DNA sequencing (*2*) is not as widely used as the Sanger method (*1*). Nevertheless, the chemical method has some advantages and can be very useful in certain situations. Sequencing can be performed from any point in the clone where a suitable restriction site occurs obviating the need for further subcloning. The sequence thus obtained can be used to design oligonucleotide primers, and further sequence can then be obtained by Sanger sequencing. The Maxam and Gilbert method is also very useful for resolving regions of DNA that yield poor results in Sanger sequencing owing to secondary structures in the DNA.

In the original Maxam and Gilbert method (*2*), a fragment of DNA is radiolabeled at one end and then partially cleaved in four different chemical reactions, each of which is specific for a particular base or type of base. This results in four populations of labeled polynucleotides. Each radiolabeled molecule extends from a fixed point (the radiolabeled end) to the site of chemical cleavage, which is determined by the DNA sequence of the original fragment. Since the cleavage is only partial, each population consists of a mixture of molecules the lengths of which are determined by the base composition of the original DNA fragment. The four reactions are electrophoresed in adjacent lanes through a polyacrylamide gel. The DNA sequence can then be determined directly from an autoradiograph of the gel. The original method has been improved over the years (*15*). Additional chemical cleavage reactions have been devised (*16*), new end-labeling techniques developed (*17,18*), and shorter, simplified protocols have been produced (*19,20*). The main advantage of chemical degradation sequencing is that sequence is obtained from the original DNA molecule and not from an enzymatic copy. It is therefore possible to analyze DNA modifications, such as methylation, and to study protein/DNA interactions. Chemical sequencing also enables the determination of the DNA sequence of synthetic oligonucleotides. However, the Sanger method is both quicker and easier to perform, and must remain the method of choice for most sequencing applications.

TEMPLATES FOR DNA SEQUENCING

Recent innovations in sequencing technology (*21*) have greatly increased the range of DNA molecules that can be utilized as templates. Sequencing reactions can now be performed directly on cosmid or λ DNA and on PCR products. However, best results, in terms of length and accuracy of the sequence obtained, are still usually achieved by using templates derived from M13 phage and from phagemid vectors. M13 phage meets the requirement of the polymerase enzyme for a single-stranded template by producing large amounts of just one strand of DNA as part of its normal replicative cycle. Double-stranded (replicative form) M13 can also be isolated, and this is used to clone the DNA fragment to be sequenced. Because of the excellent quality of sequence data obtained from M13, many researchers prefer to subclone to this vector prior to sequencing.

Sequencing reactions can also be performed directly on plasmid DNA, the double-stranded molecule being denatured prior to sequencing. Recent innovations in DNA purification techniques and the availability of improved polymerases have greatly enhanced the quality of data produced by plasmid sequencing methods (*22,23*). Cosmid clones, λ clones, and PCR products can now be sequenced directly (*24,25*) thus eliminating the need for time-consuming sub-cloning steps. The advent of cycle sequencing (*8,26–28*) has revolutionized sequencing methodology and vastly increased the range of templates that can be used.

SEQUENCING STRATEGIES

The strategy chosen will depend on the reason for sequencing and on resources available. A lot of sequencing performed is confirmatory sequencing to check the orientation or the structure of newly constructed plasmids, or to determine the sequence of mutants. This type of sequencing can easily be achieved by subcloning a restriction fragment into M13 and sequencing using the universal primer. Alternatively, a custom-designed oligonucleotide primer can be synthesized and sequencing performed without the need for any subcloning.

The determination of long tracts of unknown sequence, however, requires careful planning and the utilization of one of a variety of strategies including:

1. The shotgun approach;
2. Directed sequencing strategies; and
3. The gene-walking technique.

A random, or shotgun, approach involves subcloning random fragments of the target DNA to an appropriate vector, such as M13 (*21,29–32*). Sequences

from these recombinants are determined at random until the individual readings can be assembled into a contiguous sequence. This is achieved using a sequence assembly computer program (*33,34*). The disadvantage of this method is the redundancy in the sequence data obtained, each section of DNA being sequenced several times over; in addition, rigorous fractionation of the DNA fragments before cloning is essential to ensure appropriately sized subclones for sequencing. Since clones are then chosen at random for sequencing, shotgun methods are inefficient at completing the last few gaps in the sequence. However, the strategy benefits from making no prior assumptions about the DNA to be sequenced, such as base composition or the presence of certain restriction sites.

Directed strategies usually involve the construction of a nested set of deletions of the fragment to be sequenced. Progressive deletions of the fragment are generated with a nuclease, each deletion being approx 200–300 bp. Following deletion, the fragments are recloned into M13 or a plasmid vector adjacent to the universal primer site. The subclones are then sequenced in order of size, the sequence of each clone overlapping slightly with the one before. In this way, a large tract of contiguous sequence is determined on one gel. The disadvantage is the labor and time involved in constructing the deletions. Several methods are available for deletion construction, including the use of exonuclease III (*35*), T4 DNA polymerase (*36*), *Bal*31 (*37*), and DNAase I (*38*). It is essential to sequence both strands of the DNA, and this usually entails generating two sets of deletions.

The most straightforward and simplest method of sequencing is perhaps the gene-walking technique. It can be performed directly on plasmid clones and also on single-stranded M13. The strategy involves the initial sequencing of approx 200–400 bp of the end of a cloned fragment using the universal primer (the sequence of the other end can be achieved with the reverse primer if double-stranded template is being used). This sequence information is then used to design a new oligonucleotide primer that will provide the sequence of the next 200–400 bp and so on across the entire length of the insert. This method is the least labor intensive, since no deletion construction or generation of random clones is necessary and template DNA can be made in the one batch because the template is the same for all sequencing reactions. However, the delay involved in synthesizing a new oligonucleotide primer before the next reaction can be performed may considerably prolong the time taken to sequence a long tract of DNA. The cost of oligonucleotide synthesis may also be prohibitive.

CYCLE SEQUENCING

Cycle sequencing is a very powerful yet technically simple method of sequencing. It is a new and innovative adaptation of the Sanger dideoxy sequencing technique. The advantages over conventional sequencing

techniques are that the reactions are simpler to set up, less template is required, the quality and purity of template are not as critical, and virtually any single- or double-stranded DNA can be sequenced (including λ, cosmid, plasmid, phagemid, M13, and PCR product). In this method, a single primer is used to amplify a region of template DNA linearly using *Taq* polymerase in the presence of a mixture of dNTPs and a ddNTP. Either radioactive or fluorescent labels can be used making cycle sequencing technology as relevant to automated processes as it is to manual methods (*8,27,28*).

As in conventional dideoxy sequencing methods, cycle sequencing involves the generation of a new DNA strand from a single-stranded template, synthesis commencing at the site of an annealed primer and terminating on the incorporation of a ddNTP. The difference is that the reaction occurs not just once, but 20–30 times under the control of a thermal cycler (or PCR machine). This results in more and better sequence data from less template. The process of denaturing a double-stranded molecule is eliminated, denaturation occurring automatically in the thermal cycler. Cycle sequencing can be employed to sequence "mini-prep" plasmid DNA (*39,40*), cosmids (*24,28*), λ phage (*41,42*), and PCR products (*43,44*). Also sequence can even be achieved directly from lysates of single-plasmid- or cosmid-containing colonies, and from λ and M13 plaques (*39*). The development of cycle sequencing techniques has made a major contribution to DNA sequencing methodology, improving the reliability and efficiency of DNA sequence determination while eliminating time-consuming steps.

MULTIPLEXING AND CHEMILUMINESCENT DETECTION

The multiplex DNA sequencing strategy (*45–47*) is one of the approaches proposed recently to increase the speed and efficiency of DNA sequencing. This new strategy can be used with both the Sanger (*1*) and the Maxam and Gilbert (*2*) methods. Multiplex sequencing consists of pooling a number of samples early on in the sequencing process and processing the samples as pools. In this way, a very much larger number of samples can be handled with no more effort than that required for a small number. After electrophoresis, the DNA is transferred to a membrane, and sequence images of the individual samples are obtained by sequential hybridizations with specific labeled oligonucleotides. The additional steps required for this strategy (transfer to membrane and hybridization/washing) are more than compensated for by the reduction of work involved in template preparation, sequencing reactions, and electrophoresis.

Multiplex DNA sequencing has been used in conjunction with direct blotting electrophoresis (*48*) to facilitate transfer of the DNA to a membrane. This procedure utilizes the same electric field that separates the

polynucleotides on the gel to electroblot the DNA onto a membrane that is being moved across the bottom of the gel by a conveyor belt. Chemiluminescent detection can also be used in conjunction with multiplex DNA sequencing to visualize the image on the membrane (*49*). Several methods of labeling and detection by chemiluminescence are in use. One protocol involves 1,2-dioxetane chemiluminescence (*50*). The target DNA is immobilized on a membrane and labeled with alkaline phosphatase via an affinity system, such as the biotin/(strept)avidin or the digoxigenin system. When the dioxetane substrate is added to the membrane, the alkaline phosphatase deprotects dioxetane molecules by cleaving off the phosphate group and thus initiating the chemiluminescent reaction. The visible light produced is detected by exposure of an X-ray film.

AUTOMATION IN DNA SEQUENCING

One of the major advances in sequencing technology in recent years is the development of automated DNA sequencers. These are based on the chain-termination method and utilize fluorescent rather than radioactive labels. The fluorescent dyes can be attached to the sequencing primer, to the dNTPs, or to the terminators, and are incorporated into the DNA chain during the strand synthesis reaction mediated by a DNA polymerase (e.g., Klenow fragment of DNA polymerase I, Sequenase, or *Taq* DNA polymerase). During the electrophoresis of the newly generated DNA fragments on a polyacrylamide gel, a laser beam excites the fluorescent dyes. The emitted fluorescence is collected by detectors and the information analyzed by computer. The data are automatically converted to nucleotide sequence. A number of such instruments are now commercially available and are becoming increasingly popular (*21,51–54*).

Other aspects of the sequencing procedure that are being automated include template preparation and purification, and the sequencing reactions themselves. Robotic work stations are being developed to perform these tasks (*10*). Research is currently in progress to develop the technology of mass spectrometry for DNA sequencing (*55*). Success in this endeavor would mean that the gel electrophoresis step in DNA sequencing could be eliminated since the mass spectrometer would analyze the collection of polynucleotides produced in the Sanger procedure. Alternatively, resonance ionization spectroscopy combined with mass spectrometry could enable much faster analysis of isotopically labeled DNA bands (*55*).

COMPUTER ANALYSIS OF SEQUENCE DATA

DNA sequencing has over the years become more and more efficient, and the amount of sequence data determined has increased enormously

in recent years. The increase in the rate of accumulation of data is likely to continue at least for the next decade as sequencing technology is refined even further. The sheer volume of sequence data now available makes advanced computer methods essential to molecular biology, and a familiarity with computers and sequence analysis software is now a vital requirement for all biologists involved with DNA sequencing.

DNA sequence data are stored in several different international data banks. The EMBL (European) data library was founded in 1980 and two years later joined in a collaboration with GenBank (in Los Alamos). The Japanese DDBJ later joined forces with EMBL and GenBank, and the three data banks are now equivalent, submitted data being exchanged on a regular basis. Scientists have access to the vast amount of information stored in these repositories via on-line facilities, by obtaining copies of the data on compact disks, or via electronic mail. Several software packages exist for searching the sequence libraries for entries containing DNA similar to a query sequence. Examples include FASTA and DAP/Prosch for the mainframe computer (*56–58*). Other programs, such as XQS (*59*), permit searches by keyword, title word, or author, enabling researchers to find a sequence of interest readily. Submission of a sequence to a data bank is now a prerequisite for publication of sequence data in many journals. Release number 72 of GenBank (made available in June 1992) contains 92,160,761 bases from 71,280 entries comprising DNA sequence data from bacteria, viruses, plants, vertebrates, and invertebrates (data from GenBank).

A variety of software packages are available for analyzing sequence data (*60,61*). For extensive detailed information, *see* ref. *62*. Most packages are capable of the basic functions required by the biologist, such as translation of a DNA sequence to amino acid sequence, searching for open reading frames, drawing restriction enzyme maps and providing information on fragment sizes, manipulation of sequences by deletion or addition of segments, identification of specified targets, such as promoters or inverted repeats, and prediction of protein characteristics, such as molecular weight, hydropathy, antigenicity, and secondary structure. Some programs perform more specialized tasks, such as assembly of random fragments from shotgun sequencing projects, similarity searches, alignments of similar sequences, and phylogenetic analysis. Many programs will produce publication-quality output. Software packages are available for mainframe computers, such as the VAX, for IBM and compatible personal computers, and for the Apple Macintosh (*62*). DNA sequencing methodology and the computer analysis of the generated data are both rapidly evolving disciplines and, with the massive investment in major sequencing projects worldwide, are likely to remain at the vanguard of scientific research for the foreseeable future.

REFERENCES

1. Sanger, F., Nicklen, S., and Coulson, A. R. (1977), *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
2. Maxam, A. M. and Gilbert, W. (1977), *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
3. Hodgson, J. (1992), *Biotechnology* **10,** 760,761.
4. Oliver, S. G. et al. (1992), *Nature* **357,** 38–46.
5. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R., and Waterston, R. (1992), *Nature* **356,** 37–41.
6. Barrell, B. (1991), *FASEB J.* **5,** 40–45.
7. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989), *Molecular Cloning: a Laboratory Manual,* 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbour, NY.
8. Blakesley, R. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 26.
9. Beck, S. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin A. M., eds., Humana Press, Totowa, NJ, chs. 28–29.
10. Bankier, A. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 38.
11. Green, A. and Vaudin, M. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 25.
12. Griffin, H. G. and Griffin, A. M. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 14.
13. Durre, P. and Gerisher, U. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 13.
14. Arigoni, F. and Kaminski, (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 15.
15. Maxam, A. M. and Gilbert, W. (1980), *Methods in Enzymol.* **65,** 499–560.
16. Ambrose, B. J. B. and Pless, R. C. (1987), *Methods in Enzymol.* **152,** 522–538.
17. Volckaert, G. (1987), *Methods in Enzymol.* **155,** 231–250.
18. Eckert, R. L. (1987), *Gene* **51,** 247–254.
19. Pichersky, E. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 32.
20. Rosenthal, A. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols,* Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 33.

21. Griffin, H. G. and Griffin, A. M., eds. (1993), *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols*, Humana Press, Totowa, NJ.
22. Griffin, H. G. and Griffin, A. M. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols*, Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 18.
23. Murphy, G. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols*, Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 19.
24. Craxton, M. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols*, Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 21.
25. Gal, S. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Laboratory Protocols*, Griffin, H. G. and Griffin, A. M., eds., Humana Press, Totowa, NJ, ch. 23.
26. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K., Horn, G. T., Erlich, H. A., and Arnheim, H. (1985), *Science* **230,** 1350–1354.
27. Murray, V. (1989), *Nucleic Acids Res.* **17,** 8889.
28. Craxton, M. (1991), *Methods: A Companion to Methods in Enzymology* **3,** 20–26.
29. Bankier, A. T., Weston, K. M., and Barrell, B. G. (1987), *Methods in Enzymol.* **155,** 51–93.
30. Davison, A. J. and Scott, J. E. (1986), *J. Gen. Virol.* **67,** 1759–1816.
31. Anderson, S. (1981), *Nucleic Acids Res.* **9,** 3015.
32. Deininger, P. L. (1983), *Anal. Biochem.* **129,** 216.
33. Dolz, R. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Computer Analysis of Sequence Data*, Griffin, A. M. and Griffin, H. G., eds., Humana Press, Totowa, NJ, ch. 2.
34. Staden, R. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Computer Analysis of Sequence Data*, Griffin, A. M. and Griffin, H. G., eds., Humana Press, Totowa, NJ, ch. 17.
35. Henikoff, S. (1984), *Gene* **28,** 351–359.
36. Dale, R. M. K., McClure, B. A., and Houchins, J. P. (1985), *Plasmid* **13,** 31–40.
37. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., and Surrey, S. (1982), *Proc. Natl. Acad. Sci. USA* **79,** 4298–4302.
38. Hong, G. F. (1982), *J. Gen. Microbiol.* **158,** 539–549.
39. Krishnan, B. R., Blakesley, R. W., and Berg, D. E. (1991), *Nucleic Acids Res.* **19,** 1153.
40. Adams, S. M. and Blakesley, R. (1991), *Focus* **13,** 56–58.
41. Manfioletti, G. and Schneider, V. (1988), *Nucleic Acids Res.* **16,** 2873–2884.
42. Kim, B. S. and Jue, C. (1990), *BioTechniques* **8,** 156–159.
43. Gal, S. and Hohn, B. (1990), *Nucleic Acids Res.* **18,** 1076.
44. Green, A., Roopra, A., and Vaudin, M. (1990), *Nucleic Acids Res.* **18,** 6163.
45. Church, G. M. and Kieffer-Higgins, S. (1988), *Science* **24,** 185–188.
46. Yang, M. M. and Youvan, D. C. (1989), *Biotechnology* **7,** 576–580.
47. Chee, M. (1991), *Nucleic Acids Res.* **19,** 3301–3305.
48. Beck, S. and Pohl, F. M. (1984), *EMBO J.* **3,** 2905–2909.
49. Beck, S., O'Keeffe, T , Coull, J. M., and Koster, H. (1989), *Nucleic Acids Res.* **17,** 5115–5123.

50. Beck, S. and Koster, H. (1990), *Anal. Chem.* **62,** 2258–2270.
51. Hunkapiller, T., Kaiser, R. J., Kopp, B. F., and Hood, L. (1991), *Science* **254,** 59–67.
52. 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), *Nature* **321,** 674–679.
53. Ansorge, W., Sproat, B., Stegemann, J., and Schwager, C. (1986), *J. Biochem. Biophys. Methods* **13,** 315–323.
54. 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), *Science* **238,** 336–341.
55. Jacobson, K. B., Arlinghaus, H. F., Buchanan, M. V., Chen, C-H., Glish, G. L., Hettich, R. L., and McLuckey, S. A. (1991), *Genet. Analysis: Techniques, and Applic.* **8,** 223–229.
56. Pearson, W. R. (1993), in *Methods in Molecular Biology.* vol. *DNA Sequencing: Computer Analysis of Sequence Data,* Griffin, A. M. and Griffin, H. G., eds., Humana Press, Totowa, NJ, ch. 46.
57. Pearson, W. R. and Lipman, D. J. (1988), *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
58. Collins, J. F., Coulson, A. F. W., and Lyall, A. (1988), *Computer Appl. Bioscience* **4,** 67–71.
59. Barker, W. C., George, D. G., Hunt, L. T., and Garavelli, J. S. (1991), *Nucleic Acids Res.* **19,** 2231–2236.
60. Devereux, J., Haeberli, P., and Smithies, O. (1984), *Nucleic Acids Res.* **12,** 387–395.
61. Dear, S. and Staden, R. (1991), *Nucleic Acids Res.* **19,** 3907–3911.
62. Griffin, A. M. and Griffin, H. G., eds. (1993), *Methods in Molecular Biology.* vol. *DNA Sequencing: Computer Analysis of Sequence Data,* Humana Press, Totowa, NJ.