Essential Molecular Biology

Volume II
A Practical Approach

Edited by

T. A. BROWN

Department of Biochemistry and
Applied Molecular Biology
UMIST, Manchester M60 1QD, UK

IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

Oxford University Press, Walton Street, Oxford OX2 6DP
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Berlin Ibadan

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available
from the British Library
ISBN 0-19-963112-3 (hb)
ISBN 0-19-963113-1 (pbk)

Library of Congress Cataloging in Publication Data
Data available

Typeset by Cambrian Typesetters, Frimley, Surrey
Printed in Great Britain by
Information Press Ltd, Eynsham, Oxford

Preface

THERE are now a number of molecular biology manuals on the market and the editor of an entirely new one has a duty to explain why his contribution should be needed. My answer is that although there are some excellent handbooks for researchers who already know the basic principles of gene cloning there are very few that cater for the absolute beginner. Unfortunately, everyone is a beginner at some stage in their careers and even in an established molecular biology lab the new research student can spend a substantial amount of time not really understanding what is going on. For the experienced biologist expert in a discipline other than molecular biology, and perhaps without direct access to a tame gene cloner, guidance on how to introduce recombinant DNA techniques into his or her own research programme can be very difficult to obtain. For several years I have run a basic gene cloning course at UMIST and I have continually been impressed by the number of biochemists, botanists, geneticists, cell biologists, medics, and others who want to learn how to clone and study genes.

The contributors to *Essential Molecular Biology: A Practical Approach* were asked to write accounts that combine solid practical information with sufficient background material to ensure that the novice can understand how a technique works, what it achieves, and how to make modifications to suit personal requirements. Where appropriate the reader is also given advice on more advanced or specialized techniques. In all cases the authors have responded to the challenge and produced chapters that make concessions to the beginner without jeopardizing scientific content or practical value. I hope that the result is a handbook that will guide newcomers into molecular biology research.

The book is split into two parts. Volume I deals with the fundamental techniques needed to carry out DNA cloning experiments. The emphasis is on coming to grips with the necessary practical skills and understanding the background in sufficient detail to be able to adjust to circumstances as the project progresses. In Volume II procedures for preparing gene libraries and identifying genes are described, together with methods for studying the structure of a cloned gene and the way it is expressed in the cell. It is assumed that the basics from Volume I are now in place, but the procedures are still described in the same down to earth fashion with protocols complemented by background information and troubleshooting hints.

I must thank a number of people for their help with this book. First, I am grateful to the authors who provided the manuscripts more or less on time and were prepared in many cases to make revisions according to my requests.

62. Chang, C. T., Hain, T. C., Hutton, J. R., and Wetmur, J. G. (1974). *Biopolymers*, **13**, 1847.
63. Wetmer, J. G. (1971). *Biopolymers*, **14**, 2517.
64. Wahl, G. M., Stern, M., and Stark, G. R. (1979). *Proceedings of the National Academy of Sciences, USA*, **76**, 3683.
65. Amasino, R. M. (1986). *Analytical Biochemistry*, **152**, 304.
66. Hutton, J. R. (1977). *Nucleic Acids Research*, **4**, 3537.
67. Howley, P. M., Israel, M. F., Law, M.-F., and Martin, M. A. (1979). *Journal of Biological Chemistry*, **254**, 4876.
68. Casey, J. and Davidson, N. (1977). *Nucleic Acids Research*, **4**, 1539.
69. Britten, R. J. and Kohne, D. E. (1968). *Science*, **161**, 529.
70. Hutton, J. R. and Wetmur, J. G. (1973). *Journal of Molecular Biology*, **77**, 495.
71. Chamberlin, M. E., Galau, G. A., Britten, R. J., and Davidson, E. H. (1978). *Nucleic Acids Research*, **5**, 2073.
72. Botchan, M., Topp, W., and Sambrook, J. (1976). *Cell*, **9**, 269.
73. Denhardt, D. (1966). *Biochemical and Biophysical Research Communications*, **23**, 641.
74. Singh, L. and Jones, K. W. (1984). *Nucleic Acids Research*, **12**, 5627.
75. Johnson, D. A., Gautsch, J. W., Sportsman, J. R., and Elder, J. H. (1984) *Genetic Analytical Techniques*, **1**, 3.
76. Cos, K. H., DeLeon, D. V., Angerer, L. M., and Angerer, R. C. (1984). *Developmental Biology*, **101**, 485.
77. Gillespie, D. and Spiegelman, S. (1965). *Journal of Molecular Biology*, **12**, 829.
78. Suggs, S. V., Wallace, R. B., Hirose, T., Kiwashima, E. H., and Itakura, K. (1981). *Proceedings of the National Academy of Sciences, USA*, **78**, 6613.
79. Lathe, R. (1985). *Journal of Molecular Biology*, **183**, 1.
80. Yang, J. H., Ye, J. H., and Wallace, D. C. (1984). *Nucleic Acids Research*, **12**, 837.
81. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kiwashima, E. H., and Itakura, K. (1981). *Nucleic Acids Research*, **9**, 879.
82. Wood, W. I., Gitschier, J., Lasky, L. A., and Lawn, R. M. (1985). *Proceedings of the National Academy of Sciences, USA*, **82**, 1585.
83. Jacobs, K. A., Rudersdorf, R., Neill, S. D., Dougherty, J. P., Brown, E. L., and Fritsch, E. F. (1988). *Nucleic Acids Research*, **16**, 4637.
84. Laskey, R. A. (1984). *Review 23*, Amersham International, Amersham.
85. Sim, G. K., Kaftos, F. C., Jones, C. W., Koehler, M. D., Efstratiadis, A., and Maniatis, T. (1979). *Cell*, **18**, 1303.

6

DNA sequencing

CHRISTOPHER J. HOWE and E. SALLY WARD

1. Introduction

1.1 General principles

The 'Sanger', 'dideoxy', or 'chain-termination' method of DNA sequencing relies on the synthesis of DNA *in vitro* in the presence of chain-terminating inhibitors. Usually the template strand is provided in the form of a single-stranded molecule produced by cloning a suitable piece of DNA into the multiple cloning site of an M13 derivative, and the synthesis is primed by a short 'universal primer' oligonucleotide, complementary to a region just outside the multiple cloning site. The primer can therefore be used regardless of what has been cloned. Sequencing of double-stranded DNA (as plasmid DNA, for example, or PCR-amplified DNA) is also possible with minor modifications, but is not covered in detail here. Instead the reader is referred to more specialized manuals and papers (1, 2, 3). Usually four DNA synthesis reactions are carried out, with a different chain-terminating inhibitor added to each. The inhibitors are most commonly the 2',3'-dideoxynucleoside triphosphates of adenine, cytosine, guanine, and thymine (ddA, ddC, etc.); it is the lack of a 3'-hydroxyl group on the deoxyribose moiety that renders further elongation impossible. Under certain circumstances (for example, the formation of secondary structure by inverted repeated sequences), it may be useful to use derivatives of other bases. The synthesis is carried out in the presence of a radioactively-labelled deoxynucleoside triphosphate (usually [^{35}S]dATP) and therefore generates, for each reaction, a radioactively-labelled, nested set of oligonucleotides with identical 5'-ends (the primer) and 3'-ends all at ddA, ddG, ddC, or ddT, depending on which inhibitor was used. The products are then separated electrophoretically, on high-resolution polyacrylamide gels (which can resolve molecules differing in length by one nucleotide) and visualized by autoradiography. Sets of four gel lanes (each corresponding to one of A, G, C, or T) will thus contain a series of bands, ideally with only one at each level, and reading upwards from the bottom of the gel will give the sequence of the synthesized strand (and hence also the complementary, template strand). This information is usually fed directly into a computer, by means of a 'sonic digitizer'.

Rarely is the molecule to be sequenced short enough to be determined in a single set of reactions. Usually one of four approaches is taken. One is to clone the molecule in its entirety into M13, and use internal primers for individual sequencing reactions. This has the disadvantages that large inserts are often unstable in M13, and time and money must be invested in generating suitable primers. The second approach is to clone the entire molecule initially, and then construct a nested series of deletions, bringing successive regions close enough to the primer annealing site to be accessible for sequencing. This again may suffer from instability of the insert, and optimization of the deletion reactions may be time-consuming. The third approach is to cut the insert into smaller, defined pieces with suitable restriction enzymes and clone those pieces. This approach may be slow if the disposition of restriction sites is not convenient. The fourth approach is to fragment the DNA at random into smaller pieces (usually by sonication or DNase digestion) which can then be cloned and sequenced. This approach may also require considerable calibration at first, but can then be quite rapid. For all these strategies, computer analysis can then be used to build the total sequence from the pieces. The detailed requirements of the cloning strategies are covered elsewhere in other volumes (4). This chapter will outline the stages involved after generating M13 recombinants. Recommended suppliers of reagents and equipment are as follows:

- [α-^{35}S]dATP (400 Ci mmol^{-1}): Amersham, Cat. no. SJ 264.
- Universal sequencing primer: New England Biolabs, Cat. no. 1211 or 1212.
- Phenol (chromatographic grade), acrylamide (Electran), bisacrylamide (Electran), bromophenol blue, xylene cyanol, silanizing solution (dimethyldichlorosilane): BDH.
- dNTPs and ddNTPs: Pharmacia or Boehringer-Mannheim.
- 2'-deoxy-7-deazaguanosine 5'-triphosphate (dc7GTP or 7-deaza dGTP): Boehringer-Mannheim.
- Klenow polymerase: Boehringer-Mannheim (5 units μl^{-1}).
- Sequenase: United States Biochemical Corporation, Version 2.0 (12 units μl^{-1}).
- T7 DNA Polymerase: Pharmacia.
- Urea: Ultrapure, Gibco-BRL.
- Hamilton PB600-1 repetitive dispenser, fitted with 1710LT 100 μl gas-tight syringe with Luer tip and adaptor, Cat. no. 31330.
- Microtitre plates: Becton-Dickinson-Falcon flexible assay plates, no. 3911, round bottomed 96-well plates and lids.
- Damon IEC Centra 4X benchtop centrifuge, fitted with microtest plate head.

- Gel running apparatus: Model EV200 (20 × 20 cm), EV400 (20 × 40 cm) or EV500 (20 × 50 cm), Cambridge Electrophoresis Company.
- Plastikard 0.35 mm × 44 cm × 66 cm, Slaters (Plastikard) Ltd.
- Gel mould sealing tape: Neil Turner Ltd, Cat. no. 49LY.
- X-ray film: Kodak XAR or Fuji RX.
- GrafBar: PMS (Instruments) Ltd.

1.2 Automation

DNA sequencing is a highly repetitive process, and automation is clearly desirable. The sequencing reactions themselves and the subsequent electrophoresis and data handling are the areas that have received most attention. The Biomek 1000 Automated Laboratory Workstation produced by Beckman can be used for the former, allowing multiple and repeated pipettings to be carried out (5). A number of companies offer equipment for the latter. In general these devices substitute a fluorescent label for the radioactive one. This can be attached either to the primer or to the nucleotides that are involved in the reaction (6). The reaction products are electrophoresed on a polyacrylamide gel, past a suitably-placed detector. Use of a different label for each reaction allows all four sets of reaction products to be electrophoresed in a single lane, which helps to reduce some of the problems caused by abnormalities in gel running. The sequence information can then be fed directly into a computer. The biggest attraction of this approach is that the need for autoradiography is removed—the sequence can be read while the gel is running. Such devices are at present rather expensive, though, and the beginner is probably best advised to avoid automation, at least at first.

2. Template generation

2.1 Preparation of single-stranded DNA

The great advantage of using M13 as a cloning vector for sequencing is that particles containing single-stranded DNA are released from infected cells without lysis (see Volume I, Chapter 9). The supernatant from an infected culture should therefore contain large amounts of packaged M13 DNA, with little contamination from host nucleic acids. The phage particles are then precipitated, and the DNA extracted. The quality of subsequent sequence generation depends critically on the quality of the DNA preparation, and this is frequently the area where the beginner finds most difficulty (although often 'practice' is all that is needed). The most commonly-used technique is to infect and grow bacterial cultures in glass tubes (*Protocol 1*), although techniques have also been developed for growing material in microtitre plates (*Protocol 2*). If it is necessary to produce more DNA for a particular construct, then a μl or so of a standard single-stranded DNA preparation should be more than enough in a standard transfection experiment (see Volume I, Chapter 8).

Protocol 1. Preparation of single-stranded DNA

1. Inoculate about 20 ml of DYT medium (*Appendix 2*) with a colony of host bacteria (e.g. *E. coli* TG1; *Appendix 1*). Stand overnight at 37°C.
2. Make a 100-fold dilution into DYT, and put 1.5 ml aliquots into sterile tubes (e.g. 10 ml) with loose metal caps.
3. Transfer a single plaque into each tube, with a toothpick or pasteur pipette.
4. Shake the tubes at 300 r.p.m. for 5 to 7 h. Good aeration is necessary.
5. Transfer to Sarstedt tubes, and centrifuge for 5 min at maximum speed.
6. Decant supernatants into fresh tubes, taking great care not to dislodge any of the bacterial pellet. It is better to take less supernatant than to risk contamination. At this stage the supernatants may be stored for a day or so at 4°C. If this is done, it is advisable to add 5 µl of chloroform to each to prevent bacterial growth, and to centrifuge again before proceeding with step 7.
7. Add 200 µl of 20% (w/v) PEG 6000, 2.5 M NaCl, mix well and stand at room temp. for at least 30 min.
8. Centrifuge at maximum speed for at least 5 min. This should produce a pinhead-sized phage pellet.
9. Remove the supernatant. This can be done by pouring, turning the tubes upside down to drain, and then wiping the inside carefully with absorbent tissue. Alternatively, it can be done with a fine-tipped pipette connected to a water pump, followed by a second centrifugation to bring remaining liquid to the bottom of the tube for removal.
10. Resuspend the pellet in 100 µl TE pH 8.0 (*Appendix 2*) and add 100 µl buffer-saturated phenol. Vortex for 15 sec, and leave to stand at room temp. for at least 5 min.
11. Vortex again, and centrifuge at maximum speed for at least 5 min. This should give a sharp interface. Transfer the (upper) aqueous phase to a fresh tube, add 10 µl of 3 M sodium acetate pH 5.5 and 0.25 ml chilled ethanol.
12. Leave at −20°C overnight or at −80°C for at least 30 min to precipitate the DNA.
13. Collect by centrifugation at maximum speed for at least 10 min, decanting, washing with 1 ml chilled ethanol, decanting, and drying under vacuum.
14. Dissolve in 30 µl TE pH 8.0 (20 µl if Sequenase is to be used). Store at −20°C, and avoid repeated freeze–thawing.

Protocol 2. Preparation of template DNA in microtitre trays

1. Place 200 µl of a culture (OD_{600} = 0.2) of host cells into each well.
2. Toothpick plaques into wells. Grow with rapid shaking (300 r.p.m.) at 37°C for 4–5 h.
3. Spin the microtitre tray in a plate centrifuge for 3 min at maximum speed.
4. Transfer 150 µl of each supernatant to a new plate using a multi-pipette.
5. Add 50 µl 20% (w/v) PEG 6000, 2.5 M NaCl. Leave at room temp. for 10 min (minimum) with shaking.
6. Spin in a plate centrifuge for 10 min at maximum speed.
7. Remove the supernatant with a multi-pipette.
8. Resuspend in 75 µl TE pH 8.0 (*Appendix 2*), add an equal volume of buffer-equilibrated phenol and shake for 10 min.
9. Spin in the plate centrifuge for 10 min at maximum speed, and remove the phenol layer.
10. Add 5 µl 3 M sodium acetate pH 5.5, 130 µl ethanol to precipitate the DNA and continue as *Protocol 1*, steps 12–14.

The following points are particularly important when preparing single-stranded DNA:

(a) M13 plaques on a plate should ideally be used for producing template DNA as soon as possible after transfection, preferably within 24 h, otherwise deletion of all or part of the inserts may occur. Storage of plates at 4°C is recommended.

(b) The host strain (*Appendix 1*) should be kept on a proline selection plate (*see Appendix 2*), otherwise loss of the F′ plasmid from some cells may occur, and phage yield will be reduced.

(c) Growth conditions should be adhered to closely. Temperature control is important for good phage yields. Growing the cultures for too long may produce DNA of poor quality (degraded or contaminated because of host cell lysis).

(d) Careful removal of polyethylene glycol (PEG) and especially phenol is important, although the susceptibility of the reactions to PEG appears to vary with the particular polymerase preparation used. It should not, however, be necessary to include ether or chloroform purification steps.

2.2 Analysis of DNA preparations

It is often useful to have a rough estimate of the insert sizes of M13 DNA preparations, especially if the cloning strategy is to produce a set of nested

deletions. A procedure for this is given in *Protocol 3* (4). It may also be useful to know the orientation of inserts. This can be done by hybridization to single-stranded DNA with an insert of known orientation (4), which then alters the mobility of both molecules in an agarose gel. This is described in *Protocol 4*. Analysis of inserts is also possible using PCR (7).

Protocol 3. Size analysis of M13 DNA preparations

1. Take 20 µl of supernatant after removal of bacteria (from *Protocol 1*, step 6, or *Protocol 2*, step 4).
2. Add 1 µl 2% (w/v) SDS and 3 µl of a standard 10 × agarose gel loading buffer (Volume I, Chapter 5).
3. Run in a standard agarose gel (see Volume I, Chapter 5), with suitable known phage DNA as marker(s), and view with UV transillumination after staining with ethidium bromide.

Protocol 4. Determining insert orientation in M13 template DNA

1. Add 20 µl of supernatant to 20 µl of supernatant from a preparation of phage with an insert of known orientation.
2. Add 1 µl 2% (w/v) SDS and 4 µl of 20 × SSC (3 M NaCl, 0.3 M trisodium citrate, pH 7.0). Add a small volume of light mineral oil (which floats on the surface) to prevent evaporation, and incubate at 65°C for 1 h. Also use two preparations whose DNA is known to be complementary as controls.
3. Remove the aqueous (lower) phase, and electrophorese in an agarose gel, with individual DNAs as markers. Visualize with ethidium bromide and UV transillumination. Note that the hybridization is unlikely to go to completion, so complementarity between the inserts is indicated by the presence of three bands (one from each of the separate DNAs, and one from hybridized molecules).

2.3 Other sources of template

To sequence double-stranded plasmids, the DNA is first denatured by treatment with alkali or boiling, and then annealed to the primer (1). Several methods have also been described to generate single-stranded DNA template by PCR (1, 2). Briefly, these are (see also Chapter 7):

(a) Use of unequal molar amounts of the two PCR primers to generate an excess of one of the strands.
(b) Incorporation of biotin into one PCR primer, allowing removal of one strand by passing through a streptavidin–agarose affinity column.

(c) Use of a PCR primer with a phage promoter at the 5'-end. This promoter is then used to direct the synthesis of RNA, which is then sequenced using reverse transcriptase.

3. Sequencing reactions

3.1 Sequencing reactions with Klenow polymerase

The sequencing reactions can be carried out in capless 1.5-ml Sarstedt tubes or more conveniently in 96-well microtitre plates. Use of the latter facilitates the sequencing of large numbers of templates, as reactions for up to 24 different templates can be carried out using a single plate. Sequencing reagents are dispensed in 2 µl aliquots onto the sides of the wells, using a Hamilton repetitive dispenser, and mixed by brief centrifugation in a bench-top centrifuge fitted with a microtitre tray head. An Eppendorf 5413 centrifuge, which has a horizontal rotor, is recommended for reactions carried out in Sarstedt tubes. The reactions described in this chapter use [^{35}S]dATP, as this is safer, has a longer half-life, and, because of its lower energy of emission, causes less radiolysis of the DNA than [^{32}P]dATP and allows for higher resolution in the autoradiograph. It is important to use Klenow polymerase of high quality; we find Boehringer-Mannheim 'sequencing grade' satisfactory. The sequencing reactions are described in *Protocol 5*. The universal sequencing primer is used in these reactions, although any other suitable primer could of course be substituted.

Protocol 5. Sequencing reactions using Klenow polymerase

You should prepare the following buffers and nucleotide mixes before you start the experiment:

- TM buffer: 100 mM Tris–HCl pH 7.5, 50 mM $MgCl_2$.
- Nucleotide mixes: dissolve dNTPs and ddNTPs as individual stock solutions at 50 mM and 10 mM respectively in TE buffer pH 8.0 (*Appendix 2*). Store at −20°C.
- T, C, G, and A mixes (volumes in microlitres; store at −20°C in aliquots):

	T	C	G	A
0.5 mM dTTP	25	500	500	500
0.5 mM dCTP	500	25	500	500
0.5 mM dGTP	500	500	25	500
10 mM ddTTP	50	0	0	0
10 mM ddCTP	0	8	0	0
10 mM ddGTP	0	0	16	0
10 mM ddATP	0	0	0	1
TE pH 8.0	1000	1000	1000	500

Protocol 5. *Continued*

- Chase solution: mix and dilute dNTP stocks to give 0.5 mM dTTP, 0.5 mM dCTP, 0.5 mM dGTP, 0.5 mM dATP. Store at −20°C in aliquots.
- Formamide dye mix: 100 ml deionized formamide, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue, 10 mM EDTA pH 8.0. Store at room temp. or 4°C.

1. Annealing of template and primer. For each template, mix the following:
   - TM buffer 1 µl
   - universal sequencing primer (0.2 pmol µl^{-1}) 1 µl
   - sterile water 6 µl
2. Add 2 µl of primer mix to each of four microtitre wells (use round-bottomed wells), using a Hamilton repetitive dispenser.
3. Add 2 µl of template DNA (at about 100 ng µl^{-1}) to each well and label the wells T, C, G, and A,
4. Cover the microtitre plate with either an adhesive plate sealer or Saran wrap, ensuring that no vapour can escape; this would cause the samples to dry out.
5. Centrifuge the plate briefly to mix the primer and template.
6. Incubate the plate at 55°C for 20–30 min. Note that annealing can be carried out by adding primer mix (8 µl) to each template (8 µl) in a 1.5 ml Sarstedt tube, placing the tubes in a glass dish containing water (depth of 2 cm) which has been heated to 80°C, and allowing to cool to 40°C at room temp. This usually takes about 30 min, and may be more suitable for primers which form secondary structure because of inverted repeat sequences. The annealed template/primer mix can then, after brief centrifugation, be pipetted in 4 µl aliquots into microtitre plates. Annealed template/primer mix can be stored at −20°C for several months.
7. Centrifuge the microtitire plate briefly, to bring condensation to the bottom of the well, and remove the plate sealer or Saran wrap.
8. Dispense 2 µl of the appropriate dNTP/ddNTP mix on to the edge of the wells (e.g. 2 µl of T mix to each T well), using the repetitive dispenser.
9. Make up sufficient enzyme mix for the number of templates to be sequenced as follows (numbers correspond to volumes in microlitres). It is advisable to make slightly more enzyme mix than is required.

Number of templates	4	8	10
0.1 M dithiothreitol	4	8	10
[^{35}S]dATP	2	4	5
sterile water	24	48	60
Klenow pol (5 units µl^{-1})	2	4	5

Protocol 5. *Continued*

10. Add the Klenow polymerase last and keep on ice; dispense the enzyme mix as soon as possible after the enzyme has been added.
11. Add 2 µl of enzyme mix to each sample well, using the repetitive dispenser, and centrifuge the plate briefly. Take care not to cross-contaminate wells.
12. Incubate at room temp. (20°C) for 15 min or at 37°C for 10 min.
13. Dispense 2 µl of 0.5 mM dNTP chase mix on to the edge of each well, and mix by brief centrifugation.
14. Incubate at room temp. for 15 min at 37°C for 10 min. The plates can be stored at −20°C at this stage, for up to several months. Freezing of reactions after addition of formamide dye mix tends to cause degradation of DNA, and should be avoided if possible.
15. If the samples are to be loaded directly on to a gel, add 2 µl of formamide dye mix to the edge of each well, and centrifuge briefly to mix.
16. Incubate at 80°C for 15 min in an oven.
17. Load samples onto gels as described in *Protocol 8*.

3.2 Sequencing reactions with Sequenase and other polymerases

Sequenase and Sequenase 2.0 are the trademarks of a chemically-modified form of T7 DNA polymerase which incorporates dNTPs and ddNTPs at a more uniform rate and with a higher degree of processivity than Klenow polymerase. Use of Klenow polymerase results in variations in intensities of bands on sequencing gels, because of different affinities of this enzyme for its substrates. The most marked example of this is the observation of a weak first C in a run of Cs, sometimes making the first C unreadable. In addition, sequence-specific pile-ups (often co-migrating with A bands in the sequence PuCA) are often observed because of pausing of the polymerase. The use of Sequenase overcomes these problems, resulting in a more even band intensity. It is therefore of particular use when automatic gel readers are being employed. United States Biochemical Corporation (USB) is currently the recommended supplier of this enzyme, and has produced an improved version (Sequenase 2.0). Version 2.0 is a genetically-modified form of the earlier version of Sequenase, and in contrast to the latter, has no detectable exonuclease activity. Pharmacia also markets an unmodified form of the enzyme. Both Pharmacia and USB supply the enzyme in convenient kits containing nucleotide mixes, primers and buffers, together with a booklet

containing detailed reaction protocols. The recommended reaction procedure is rather different from that using Klenow polymerase, in that annealing is followed by a labelling/extension step, during which no ddNTPs are present. This step is carried out using [^{35}S]dATP and limiting concentrations of the other three dNTPs. The length and temperature of the labelling/extension step, and concentration of dNTPs and template DNA are critical determinants of the length of extension from the annealed primer. It is therefore important to tailor the labelling/extension conditions to obtain sequence information for the region being analysed. For example, too short a labelling/extension step will result in strong readable bands at the bottom of the gel near the primer-annealing site, which soon become unreadable further up. The labelling/extension reaction is followed by a termination step, in which all four dNTPs and a relatively high concentration of a specific ddNTP are added to each reaction, However, to allow the use of a 2 µl Hamilton repetitive dispenser and microtitre plates, the reagent volumes quoted in the USB protocol need modification. These modified conditions (A. Bankier and C. Brown, personal communication) are shown in *Protocol 6*. *Taq* (*Thermus aquaticus*) polymerase may also be useful, especially when high reaction temperatures are needed (when sequencing GC-rich regions) and in sequencing of PCR-amplified material.

Protocol 6. Sequencing reactions using Sequenase

You should have the following buffers and nucleotide mixes ready before you start the experiment:

- 5 × Sequenase buffer: 200 mM Tris–HCl pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl.
- Labelling/extension mix: 2 µM dTTP, 2 µM dCTP, 2 µM dGTP. Store at −20°C in aliquots.
- Sequenase termination mixes: each mix contains all four dNTPs at 150 µM, and the appropriate ddNTP at a concentration of 15 µM. Store at −20°C in aliquots.
- Formamide dye mix; as in *Protocol 5*.

1. Annealing of template and primer. For each template, mix the following:
   - 5 × Sequenase buffer — 2 µl
   - universal sequencing primer (0.5 pmol µl^{-1}) — 1 µl
   - sterile water — 6 µl
2. Add 2 µl of primer mix to each microtitre well, using a Hamilton repetitive dispenser.

Protocol 6. *Continued*

3. Add 2 µl of template DNA (at about 150 ng µl^{-1}) to the T, C, G, and A wells. Note the different quantities of template; usually 50% more template is used in Sequenase reactions. This is achieved by resuspending the template preparation in a smaller volume of TE at the final steps of *Protocols 1* and *2*.
4. Anneal as in *Protocol 5*, steps 4–7.
5. Using the repetitive dispenser, add 2 µl of labelling/extension mix, made up as follows (volumes in microlitres):

Number of templates	4	8	10
0.1 M dithiothreitol	4	8	10
1 × labelling/extension mix	1	2	2.5
[^{35}S]dATP	2	4	5
sterile water	24	48	60
Sequenase	1	2	2.5

6. Mix by brief centrifugation and incubate at room temp. (20°C) for 5 min. Note that these are suitable labelling/extension conditions for reading a sequence from the primer site to about 200 bases distal to the primer. Sequence further from the primer site can be generated by increasing the amount of labelling/extension mix and/or the labelling/extension time. In addition, the concentration of template DNA in the labelling/extension reaction will affect the extension.
7. Dispense 2 µl of appropriate termination mix pre-warmed to 37°C on to the edge of each well, using the repetitive dispenser.
8. Mix by brief centrifugation and incubate at 37°C for 5 min.
9. Add 2 µl of stop (formamide dye mix) solution to the edge of each well, and centrifuge briefly to mix.
10. Incubate in an oven at 80°C for 15 min.
11. Load the samples on to a gel as described in *Protocol 8*.

3.3 Use of dGTP analogues during sequencing reactions

GC-rich regions tend to form hairpin loops which result in co-migration of the bands on a gel to form a 'compression', often making the sequence unreadable in this region. To resolve this problem the gel can be run in such a way as to create a more denaturing environment. This can be done by electrophoresing

at a higher temperature or including formamide in the gel mix. Alternatively, the problem can be overcome with dGTP analogues which do not form stable GC base-pairs. Two analogues which can be used are dITP and 7-deaza dGTP, and nucleotide mixes for Klenow polymerase and Sequenase reactions using one or both of these instead of dGTP are shown in *Tables 1* and *2*. Use of these analogues tends to result in a sequence of poor quality, however, and therefore is recommended only when alterations in gel running conditions cannot resolve the problem. Moreover, the sequence around the compression can generally be resolved by sequencing the complementary strand, which is of course necessary in any sequencing project.

4. Gel electrophoresis

4.1 Denaturing polyacrylamide gel electrophoresis

High-resolution polyacrylamide gels will separate strands which differ in length by one base, up to a limit, and the amount of readable sequence obtainable from a gel run is dependent on the length of the gel and the running time. The other limitations on the amount of sequence information

Table 1. Nucleotide mixes for dITP and 7-deaza dGTP sequencing with Klenow polymerase

(A) Deoxyinosine triphosphate (dITP) mixes (volumes in µl)*

	T	C	G	A
0.5 mM dTTP	25	500	500	500
0.5 mM dCTP	500	25	50	500
2.0 mM dITP	500	500	25	500
10 mM ddTTP	50	0	0	0
10 mM ddCTP	0	8	0	0
10 mM ddGTP	0	0	2	0
10 mM ddATP	0	0	0	1
TE pH 8.0	1000	1000	1000	500

(B) 7-deaza dGTP mixes (volumes in µl)*

	T	C	G	A
0.5 mM dTTP	25	500	500	500
0.5 mM dCTP	500	25	500	500
0.5 mM 7-deaza dGTP	500	500	25	500
10 mM ddTTP	50	0	0	0
10 mM ddCTP	0	8	0	0
10 mM ddGTP	0	0	16	0
10 mM ddATP	0	0	0	1
TE pH 8.0	1000	1000	1000	500

* Store at −20°C in aliquots.

Table 2. Nucleotide mixes for dITP sequencing with Sequenase

(A) Labelling/extension mix*

- dTTP 2 µM
- dCTP 2 µM
- dITP 4 µM

(B) Termination mixes (concentrations in µM)*

	T	C	G	A
dTTP	150	150	150	150
dCTP	150	150	150	150
dITP	300	300	300	300
dATP	150	150	150	150
ddTTP	15	0	0	0
ddCTP	0	15	0	0
ddGTP	0	0	3	0
ddATP	0	0	0	15

* Store at −20°C in aliquots.

obtainable from a particular sequence reaction are obviously the reaction conditions, quality of enzyme, template, etc. It is critical to run the sequencing gels under denaturing conditions, as described in the following sections.

The gels used for the analysis of the products of sequencing reactions are usually 20 cm wide and between 40 and 50 cm long; the precise size depends on the gel tank. Gel moulds consist of two glass plates sandwiching spacers, which are placed down each long edge. One of the plates has a notch cut out of the top to allow efficient contact of the gel-running buffer with the gel. It is important that gels are as thin as is practicable, as this will maximize resolution, although it will also lead to a reduction in the volume of sample that can be loaded. Plastikard (polystyrene) spacers of 0.35 mm thickness are recommended, and the gel comb is cut from the same material. It is important that the gel teeth are machine cut to produce wells which are flat-bottomed, otherwise poor-quality gels result. The number of gel wells per comb is a matter of personal preference, but if 96-well microtitre plates are being used a 50-teeth comb (wells 2 mm wide and 3 mm deep) is convenient, so that two gels are sufficient to analyse 96 reactions. If fewer samples are being analysed, combs with fewer teeth are easier to use. Teflon combs and spacers, and 'sharkstooth' combs are supplied by some companies.

The usable width of a 20-cm-wide gel is generally about 16 cm, as considerable 'smiling' (slower running at the edges of the gel) occurs because of unequal heat distribution across the width of the gel, if wells are loaded too close to the edges. Excessive smiling causing difficulty during gel reading, particularly if this is automated.

4.1.1 Preparation and pouring of the polyacrylamide gel

Protocol 7 gives the methods for casting thin denaturing polyacrylamide gels. Several points should be emphasized. The gel plates should be as clean and dust-free as possible, otherwise troublesome air bubbles will result during pouring. It is sometimes possible to dislodge these by placing the gel in an almost vertical position after pouring, and tapping the plates. Alternatively, bubbles may be dislodged by sliding a strip of X-ray film between the plates. Bubbles and particulate matter in the gel matrix cause streaking of the samples during gel running.

Gels should be left to polymerize for at least 30 min before running, and can be left overnight at room temperature provided the top of the gel is covered with dampened Kimwipes and Saran wrap to prevent drying out. It is important to allow the gel to polymerize for a sufficient length of time: if the comb is removed before polymerization is complete, poor wells are produced which result in a sequence of low quality.

Protocol 7. Preparation of a denaturing polyacrylamide gel

You will require the following buffers and reagents:

- Ammonium persulphate: make up at 25% (w/v) in deionized water. Can be stored at 4°C for several months in the dark.
- TEMED = *N,N,N',N'*-tetramethyl-1,2-diaminoethane.
- 10 × TBE: dissolve 108 g Tris-base + 55 g boric acid + 9.3 g EDTA in deionized water and make up to 1 litre. Store at room temp.; after a few weeks a white precipitate may form. Discard when this occurs.
- 40% (w/v) acrylamide: dissolve 380 g acrylamide + 20 g *N,N'*-methylene bisacrylamide in deionized water and make up to 1 litre. Deionize by stirring with Amberlite MB1 (2%, w/v) for 5 min and filter through a Whatman 1 filter disc. Stable at 4°C for several months.
- 6% (or 4%) (w/v) denaturing acrylamide gel mix: 460 g urea, 150 ml 40% (w/v) acrylamide (or 100 ml for 4%), 100 ml 10 × TBE made up to 1 litre with deionized water. Store at 4°C. Stable for several months.
- 0.5 × TBE denaturing acrylamide mix (buffer gradient gels): 460 g urea, 50 ml 10 × TBE, 150 ml 40% (w/v) acrylamide made up to 1 litre with deionized water. Store at 4°C. Stable for several months.
- 5 × TBE denaturing acrylamide mix (buffer gradient gels): 92 g urea, 30 ml 40% (w/v) acrylamide, 100 ml 10 × TBE, 10 mg bromophenol blue (this facilitates the visualization of the gradient as it is mixed and poured) made up to 200 ml with deionized water. Store at 4°C. Stable for up to a month.

Protocol 7. *Continued*

1. Clean the glass plates thoroughly in warm water. It is advisable not to use abrasives or detergents to clean the plates, and this should not be necessary if the plates are washed immediately after use.
2. In a fume cupboard, wipe a few millilitres of silanizing solution over the surface of the notched plate with Kimwipes, on the side which will contact the gel. Leave to air-dry.
3. Wipe both plates with 96% (v/v) ethanol, and polish thoroughly, taking care to remove dust from the plate surfaces.
4. Assemble the gel mould with the two spacers, and use gel-sealing tape to seal around the bottom and sides of the mould. Ensure that sealing is complete.
5. Allow sufficient 4% or 6% gel mix to warm up to room temp. For a 40 cm × 0.35 mm × 20 cm gel mould use 40 ml; for a 50-cm-long gel use 50 ml.
6. For 40 ml of gel solution add 70 µl 25% (w/v) ammonium persulphate and 70 µl TEMED. Mix and then pour into the gel mould using a 50-ml disposable syringe. Start pouring the gel with the mould at about 45° to the horizontal, and gradually decrease this angle as the mould fills up with gel mix. Try to keep the flow of gel mix between the plates continuous, as this will minimize the risk of air bubbles. Keep any remaining gel mix on ice, for step 7.
7. Insert the gel comb to a depth of about 0.5 cm, clamp the plates with foldback clips directly over the spacers and leave the gel to polymerize for at least 30 min. Check that the top of the gel level does not drop, and top up with the remaining gel mix if necessary.

4.1.2 Gel apparatus

Many types of gel apparatus are currently available, and some of these have inbuilt passive or active temperature control systems. These temperature control systems are designed to dissipate the temperature gradient which develops during gel running and causes smiling. Passive temperature control systems have a large volume of buffer in contact with one of the gel plates, acting as a heat sink. Active control, on the other hand, involves the pumping of buffer from a reservoir which is temperature-controlled around a chamber which contacts one side of the gel mould. These systems have the advantage of allowing gels to be run at precisely-controlled temperatures so that, for example, compressions can be melted and resolved. However, the extra expense of these controlled systems generally outweighs their advantages. The simple design shown in *Figure 1* is much cheaper and is adequate for most purposes. In addition, an aluminium sheet can be placed in direct contact with the outer glass plate of the gel mould during gel running, to dissipate the

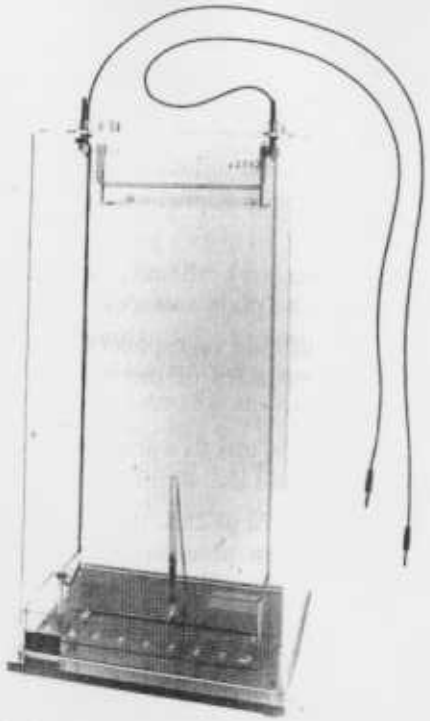

Figure 1. A simple vertical gel electrophoresis apparatus. The apparatus has a safety cover, but no inbuilt cooling system; this type of system is perfectly adequate for most sequencing purposes.

temperature gradient and prevent excessive smiling. These systems have the added advantage that if the upper buffer trough and electrode is designed to be supported by the gel mould, different lengths of gel can be readily accommodated.

4.1.3 Gel loading and running

Protocol 8 describes the preparation of gels for running, once poured, and the steps taken to load the denatured samples. As soon as the gel comb is removed urea leaches out of the gel into the sample wells. The urea should be flushed out of the gel wells immediately before loading, as its continued presence will result in diffuse bands. Gel samples should be loaded as rapidly as possible, to minimize renaturation and diffusion of the samples in the gel wells. To avoid cross-contamination of the samples the gel loader should be thoroughly rinsed in the lower buffer chamber between each application. The ideal gel loader, although less desirable on safety grounds, consists of a length of rubber tubing connected to a mouthpiece at one end and a piece of narrow bore FEP tubing which has been pulled to a fine point in a flame. The tubing

must be narrow enough to fit between the gel plates, but not so narrow as to make gel loading difficult. Care must obviously be taken during loading to avoid blowing air bubbles through the samples in the gel wells. This will not only lose the sample being loaded, but will also result in cross-contamination of the other sample wells. Loading of sequencing gels usually requires some practice, and the beginner is advised to start with gel combs with wider teeth than those on the 50-well combs used by more experienced sequencers.

Protocol 8. Preparation of a sequencing gel for running

Prepare the gel for running a few minutes before loading.

1. Remove or slit with a razor blade the gel-sealing tape along the bottom of the cast gel.
2. Place the gel mould in the gel running tank, and clamp securely into place. It may be necessary to seal around the upper rubber seal of the apparatus with silicone wax to prevent buffer leaking.
3. Pour 1 × TBE buffer (usually about 400 ml per chamber) into the upper and lower buffer chambers.
4. Remove the gel comb, and *immediately* flush out the gel wells with a pasteur pipette and the buffer in the upper chamber.
5. The samples should be heat-denatured by heating at 80°C for 15 min, and the sample volume reduced to about 3 µl.
6. Immediately prior to loading, flush out the sample wells again to remove any residual urea. Suck the sample into the narrow-bore pipette, and apply about 90% of the sample into the gel well, by gentle blowing. Place the end of the capillary close to the bottom of the well and do not attempt to load all of the sample, as this will probably introduce air bubbles. Do not load samples into the wells to a depth greater than half the width of the well, as this will result in loss of resolution.
7. Rinse the capillary, and load the next sample, proceeding as quickly as possible.
8. When gel loading is complete, place the cover over the gel (or safety cabinet) and turn the power supply on. Run the gel at constant power, for a 50-cm-long gel run at 35–40 watts (current = 25–32 mA; voltage = 1.3–1.5 kV) for about 2.5 h, and for 40 cm gels run at the same power for 1.5 h. By this time, the bromophenol blue marker dye should reach the end of the gel.

4.1.4 Length of run

The rate-limiting part of a sequencing project is generally the running of the gels, and it is therefore desirable to obtain the maximum amount of sequence

from a set of reactions and the subsequent gel. The information obtained can be maximized in one of several ways, namely, multiple loading of samples, or the use of either wedge or buffer gradient gels. With good resolution, a 40 cm 6% gel will yield about 200 nucleotides of sequence. If the gel is 50 cm long, this can be increased to 250 nucleotides. Running longer gels for extended periods will not, however, increase the amount of readable sequence proportionally, and moreover the long gels become difficult to handle. In addition, sequence information in proximity to the priming site is lost from the bottom of the gel during prolonged runs. This problem can be overcome in part by multiple loading, that is, loading of portions of the same sample at time intervals onto the same or different gels. For example, samples can be run on gels for 2, 4, and 6 h by loading at staggered intervals; this will generate up to 450 bases of readable sequence. During prolonged runs it is advisable to change the running buffer halfway through. In addition, the sample can be run on 4% gels for prolonged periods to obtain readable sequence further from the priming site. It is unlikely, however, that more than 450 bases of accurate sequence can be obtained using this approach, as the gel resolving power becomes limiting, making the determination of the number of nucleotides in a run of multiple bands, for example, particularly difficult.

4.2 Voltage gradient gels

Running samples on a linear 6% (or 4%) gel with a uniform field strength produces a wide separation of bands near the bottom of the gel. The bands gradually become closer as one reads further from the primer annealing site. This wastes gel space, and the ideal would be to have an even spacing of bands all the way up the gel. A step towards this can be taken by having a gradient of field strength, decreasing down the gel. Thus when smaller DNA molecules move further down the gel, they experience less driving force. Such a gradient can be set up in one of two ways, using wedge or buffer gradient gels (8, 9). The preparation and running conditions for these gels are given in *Protocols 9* and *10*.

4.2.1 Wedge gels

Wedge gradient gels are produced with wedge-shaped spacers. Since the electrical resistance of a TBE/polyacrylamide gel decreases with cross-sectional area, a linear increase in the gel thickness from top to bottom results in a decrease in voltage gradient towards the bottom of the gel. The limit on the maximum thickness of the wedge is 1.5 mm, as the required drying time and the band fuzziness both increase and become limiting as the thickness is increased further.

Protocol 9. Preparation and running of wedge gels

1. Use polystyrene cement fluid to fix two 3 cm Plastikard strips (one on either side) on one end of each of two spacers.
2. Assemble the gel mould and pour the gel as described in *Protocol 7*. Clamp the gel, once poured, at the top and directly over the thick end of the spacer.
3. Load, run and process the gel as described in *Protocols 8* and *11*. Remember that the gel will take longer to fix (about 30 min) and dry down (about 45 min).

4.2.2 Buffer gradient gels

Buffer gradient gels are an attractive alternative, as their use overcomes the problems encountered in drying down wedge-shaped gels. They are, however, slightly more difficult to pour than wedge gels. The gradient is produced by slight mixing of two gel mixes of different TBE strengths in a pipette, and then pipetting the gradient between the gel plates of the mould. The gradient is limited to the lower part of the gel mould only, and its length and steepness can be altered by changing the volumes and concentrations of the different TBE mixes used. The degree of mixing of the gradient solutions in the pipette can also be varied to alter the gradient properties. It is possible to obtain 250 bases of sequence from a 40 cm long buffer gradient gel; use of a 50 cm long gel extends this to about 300 bases. It is advisable, however, not to run buffer gradient gels for longer than is needed to run the bromophenol blue dye to the bottom of the gel, as the gels become very hot during prolonged runs, tending to cause cracking of the plates.

Protocol 10. Preparation and running of buffer gradient gels

1. Assemble the gel mould as described in *Protocol 7*, steps 1–4.
2. For a 20 cm × 50 cm × 0.35 mm gel, allow 50 ml of 0.5 × TBE gel mix and 7 ml of 5.0 × TBE gel mix to reach room temp. in separate containers.
3. Add 100 µl each of 25% (w/v) ammonium persulphate and TEMED to the 0.5 × TBE mix, and 14 µl each to the 5.0 × TBE mix.
4. Draw 40 ml of 0.5 × TBE mix into a 50-ml disposable syringe, and set aside.
5. Fit a pipette controller on to a 10 ml pipette, and suck up 6 ml 0.5 × TBE mix, followed by 6 ml 5.0 × TBE mix, to form two layers.
6. Introduce a few air bubbles to form a crude gradient in the pipette. The amount of mixing can be altered to produce gradients of differing

Protocol 10. *Continued*

steepness. Addition of bromophenol blue to the 5.0 × TBE gel mix (*Protocol 7*) allows the degree of mixing to be visualized.

7. Transfer the gradient mix to the gel mould by pipetting it down one edge of the mould. Hold the mould at an angle, and keep the flow continuous. The gradient mix can be poured down the centre of the mould; this will result in a more even gradient, but is more difficult to do.
8. When all the gradient mix has been poured between the plates, lower the plates to the horizontal to arrest the flow.
9. Holding the mould at an angle to the horizontal, pour in the 0.5 × TBE mix, keeping the flow continuous with that of the gradient mix.
10. Lower the mould to the horizontal, when filled, and insert the gel comb to a depth of about 0.5 cm.
11. Clamp the plates together, and continue gel loading, running, and autoradiography as in *Protocols 8* and *11*, except that a 40 cm gel will take about 2.5 h, and a 50 cm gel about 3.5 h, to run.

4.3 Autoradiography

Due to the low energy emission of [^{35}S]dATP it is important to dry sequencing gels to minimize their thickness prior to exposure to X-ray film, otherwise the signal is severely quenched. In contrast, if [^{32}P]dATP is being used, gel drying is preferable but not essential. Close contact between the gel and X-ray film is necessary if strong, sharp bands are to be obtained. The procedure for preparing and drying gels for autoradiography is as follows:

Protocol 11. Preparation of gels for autoradiography

1. When electrophoresis is complete, disconnect the power supply and discard the TBE buffer (the lower buffer, in particular, will contain relatively high levels of radioactivity, and should be disposed of appropriately).
2. Remove the gel mould from the chamber, and remove sealing tape and spacers.
3. Prise the gel plates apart, using either scissors or a spatula, with the silanized notched plate uppermost (the gel should not stick to this plate, although, despite silanization, this sometimes occurs).
4. Place the plate with the gel uppermost in a suitable shallow tank (we routinely use seed trays), and gently cover with 10% (v/v) acetic acid as fixer. Leave the gel in fixer for at least 15 min, and then cover the gel with plastic mesh (e.g. greenhouse shading), followed by the notched plate, and carefully lift from the tank. It is essential to fix the gel for at

Protocol 11. *Continued*

least 15 min, to allow the urea to leach out from the gel; unless this occurs, it is not possible to dry the gel down properly. If changes in the refractive index of the fixer are observed when the gel is gently agitated, then there is still urea in the gel.

5. Remove the notched plate and mesh and hold the plate and gel at about 20° to the horizontal, so that residual fixer drains off (take care that the gel does not slide off the plate!).
6. Remove excess liquid with Kimwipes, and then lay a piece of Whatman 3MM paper, cut to a suitable size, over the gel, and slowly peel back. The gel should stick to the paper.
7. Cover the gel and paper with Saran wrap and dry the gel under vacuum at 80°C for 15–30 min. It is essential that the gel is dried properly to prevent quenching of the radiation and the gel sticking to the X-ray film. Alternatively, if [^{32}P]dATP is being used, the gel can be exposed to X-ray film wet, with a piece of Saran wrap between the gel and film. (Note that domestic cling-film is too permeable for this.)
8. Remove the Saran wrap from the dried gel, and place in a cassette next to the X-ray film.
9. Expose for 24 h or longer, at room temp.
10. Develop the film using the method recommended by the supplier.

5. Data analysis

5.1 Fault finding

The autoradiograph usually provides the first indication as to whether or not the previous steps have been successful, and therefore yields a lot of useful information for trouble-shooting. The most common problems involve the presence of extra bands ('artefact' bands), faint bands or diffuse bands (10). Causes are suggested below. In addition, certain sequence-specific artefacts are often encountered.

5.1.1 Artefact bands in one track only

These are usually caused by having an inappropriate ratio of dNTP to ddNTP, or contamination or degradation of a mix. Sometimes (especially in the C track when using Klenow enzyme) they may be due to use of a stale dithiothreitol solution. Make up fresh solutions.

5.1.2 Artefact bands in all tracks

This is usually indicative of a poor-quality template. This may be due to contamination (e.g. with PEG, NaCl, phenol, or cell debris caused by growing the cells for too long before harvesting the single-stranded DNA). It

may also be caused by degradation of the DNA (again perhaps through growing the cells for too long, excessive freeze–thawing of the template or freezing the sequencing reaction products after addition of formamide). Other, less frequent, causes are the presence of excess primer during annealing, use of poor quality polymerase, cross-contamination of templates (perhaps through failure to pick single plaques earlier) or mixes. If DNA known to be of good quality is available, then that should be used to see if template quality is a problem. Increasing the polymerase concentration may help, but preparation of fresh template is usually the best answer.

5.1.3 Bands faint

This may be caused by a low yield in the DNA preparation. Check that the cells used for growth of phage have come from a suitable selection plate, and that the temperature and other conditions used were correct. Freeze–thawing of template or primer, or use of faulty nucleotide mixes may also be responsible. If this is suspected, replace these solutions. Obviously, use of old radiolabel will cause bands to be faint. Complete absence of bands may be due to deletions within the recombinant phage leading to loss of the primer-annealing site.

5.1.4 Bands diffuse

This usually indicates a problem with the electrophoresis or preparation for autoradiography. The most frequent causes are overheating or prolonged heating of the samples before loading, overloading of the samples (the thinner the gel the more likely this is to be a problem), failure to wash the urea out of the wells before loading or out of the gel before drying down, stale or poor quality electrophoresis reagents (check that the 10 × TBE has not produced a white precipitate), dirty gel plates, failure to allow the gel to polymerize properly before electrophoresis, and gel overheating during electrophoresis or drying down.

5.1.5 Sequence-specific artefacts

The following are among those noted more commonly:

(a) A bands: in a run of As, the bottom band is often stronger than the others.

(b) C bands: where two or more Cs are adjacent, the lowest is generally much weaker than the next one. This can be so severe that the lowest C may not be visible at all, and this may also occur with single C bands. It is generally much worse with Klenow polymerase than with Sequenase.

(c) G bands: these may be weak in the sequence TG.

(d) TGCC: this sequence may cause an artefact band in the C track at or between the levels of the T and G bands.

(e) GCA: this sequence may cause an artefactual T or C at the level of the A band.

(f) Compressions are probably caused by the formation of hairpin loops during electrophoresis. They cause bands to run very close together, sometimes superimposed, often with increased spacing in the region immediately above. Addition of formamide to the gel may help this problem, as may modification of the sequencing reactions (see Section 3.3). Use of thermostatted plates may also help.

(g) Pile-ups are also known as 'walls', and are recognized as strong stops in all four tracks. Carrying out the reactions at higher temperatures with *Taq* polymerase may help, and sequencing on the opposite strand usually also resolves any ambiguity. Sequencing both strands should in any case be regarded as essential to any sequencing project.

5.2 Gel transcription with a sonic digitizer

Transcribing sequence from a gel by hand and then typing the information into a computer is both tedious and unreliable. The easiest solution is to use a GrafBar sonic digitizer, which is attached to the top of a light box. The autoradiograph is placed perpendicular to the GrafBar, and a pen-like device is moved up the autoradiograph, with the tip being pressed on to each base in turn. Pressing the pen causes a signal to be emitted, which is detected by two sensors on the GrafBar. The position of the pen is then computed, which indicates which lane has been signalled, and thus the order of the bases in the sequence. At the same time, a note is emitted, to confirm audibly which base has been called. The sequence information is fed directly into a suitable computer attached to the digitizer. If the lanes in the autoradiograph are not sufficiently straight (as a consequence perhaps of a badly-poured buffer gradient or serious smiling) it may be necessary to redefine the working area of the gel. This will usually be indicated by the program.

5.3 Computer systems available for sequence analysis

The computer systems available are constantly evolving, and it is probably inappropriate to go into much detail in a short review such as this. More detailed descriptions are available elsewhere: we will describe some of the systems available (this section), and indicate the main kinds of program needed (next section) for data handling (11, 12).

Very often, a small computer (e.g. Apple Macintosh, IBM PC, or similar) in an individual lab is used as a terminal for a much larger machine, such as a VAX. Access to such a machine is essential for operations such as database searching. Larger individual machines might include a DEC Micro VAX or a SUN (Unix) workstation. If access to a VAX is available, then this with (for example) a Macintosh as a terminal represents perhaps the optimal combination of machines, although it should be emphasized that this is to a large extent a matter of personal preference and that for some applications

other combinations may be more desirable (some complex evolutionary tree-building may be better performed on more powerful machines, for example). A wide range of programs is available for the above and other machines, and these are continually being modified, and moved to other operating systems. Probably the largest and most widely-used general package at the present is the STADEN system, which is available for IBM PCs, VAX/VMS, and Unix machines. The GCG package (VAX/VMS, Unix) is also widely used.

5.4 Sequence analysis

The first stage is to use the sonic digitizer to enter the sequence into a computer. A suitable program is required to run the digitizer. These include READGEL (IBM PC), GELIN (STADEN) and GelEnter (GCG). It is important to check that vector sequence has not been read in by mistake (for example by reading all the way through a short insert), and many packages have facilities for removing gel readings that are 'contaminated' with vector sequence. The vector region should be edited out, and the remaining sequence used. The next step is usually the assembly of individual readings into (ideally) one unambiguous sequence covering the entire region in question. It would be inappropriate to go into great detail here as documentation should of course be obtained with the programs. In principle, each gel reading is systematically compared (using both strands) with the others. If the computer finds overlap between two sequences, they are placed together in a 'contig'. If two contigs overlap, they can also be joined to form a larger contig and so on. Ideally, it should be possible to join all the individual gel readings into one large contig, but a number of factors may prevent this. One is that not all the sequence may have been determined. It may be necessary later to identify specific restriction fragments which will be able to link contigs. These can then be cloned and sequenced as appropriate. The next problem is that a piece of sequence may be of poor quality, and the computer does not regard it as being sufficiently similar to other sequences to form a contig. Most programs allow some flexibility, for example in allowing a certain number or percentage of gaps or mismatches between two sequences. In addition, when uncertainty exists in a gel reading, it is usually possible to enter an 'ambiguity' code letter to indicate this. For example, in the STADEN system, 'D' means 'C or CC', 'R' means 'A or G', and there are other codes to cover all eventualities. The program can then make allowances for these ambiguities in generating contigs. Because there may be differences between individual gel readings, it is not usually possible to derive one single overall sequence on the first run. A 'best fit' or consensus is calculated, and discrepancies between individual readings can then be compared with the gels to resolve disagreements. This process can be very tedious, and for this reason poor quality sequence can be a hindrance rather than a help. Because of the need to go back and consult gels, it is very important to give systematic

names to individual readings that are entered into the computer, otherwise identifying the gels involved becomes a very difficult task! Once a consensus sequence has been derived, and discrepancies resolved, analysis of the data can be carried out. The following manipulations can usually be carried out quite readily.

5.4.1 Deriving a restriction map

A computer file containing lists of restriction enzyme recognition sites is compared with the sequence, and a listing of sites generated. This need not be confined to restriction sites, of course. Promoter motifs and ribosome binding sites, for example, can be screened for in similar fashion.

5.4.2 Secondary structures and repeats

The sequence can be analysed for secondary structures, such as hairpin loops. Hairpin loops will be inverted repeat sequences; it is also possible to screen for direct repeats. More sophisticated programs can screen for complicated structures such as tRNA coding regions, and allow for the presence of introns, mismatches and so on.

5.4.3 Coding regions

Coding regions have a number of features that can be used in analysing a sequence. There should be a long stretch of sequence with at least one reading frame devoid of termination codons. An AUG codon might also be expected, although other codons can occasionally be used for initiation, and in any case the presence of introns may complicate this. Coding regions also have a non-random disposition of bases, so this, combined with the lack of a termination codon would be particularly compelling evidence of a coding function. Translation programs, using either the standard genetic code or a specified variant of it, will convert the DNA sequence into the corresponding amino acid sequence. Just as with the screening of DNA for particular motifs, protein sequence can be screened for important features, such as glycosylation sites, processing sites and so on.

5.4.4 Database searching and homology displays

A number of databases, of both DNA and amino acid sequences, are maintained and distributed widely. Many journals make it a condition of acceptance of a manuscript that evidence is provided that any sequence information has been submitted to a database (consult the 'Instructions to Authors' for this). Screening large databases will of course require a quite powerful machine, but may well be very rewarding. Homologies between sequences are often most easily displayed in the form of a DIAGON plot. This is a two-dimensional representation, where one sequence is placed along one axis, and the other is placed perpendicular to it. Where the same residue occurs in both sequences, a dot is placed at the appropriate co-ordinates on

the plot. A region of homology between two sequences will therefore be indicated by a diagonal line. In practice it is much better to compare not individual residues, but a window moving down the sequences. The length of the window and the accuracy of match required can be selected and will affect the background, the length of any homologies and so on. When using these techniques, it is advisable to experiment with a wide range of parameters to avoid missing anything.

5.4.5 Hydrophobicity and secondary structure

The degree of hydrophobicity, and the likelihood of forming particular structures (alpha-helix, beta-pleated sheet, etc.) can be plotted along the length of the protein, and predictions made about the structure of the protein (12). Integral membrane proteins, for example, often contain stretches of membrane-spanning hydrophobic alpha-helices, separated by hydrophilic regions. However, the algorithms should be regarded with a degree of scepticism since (contrary to the indications given by much of the literature) they are far from being completely reliable. The importance of testing all these predictions experimentally cannot be over-emphasized!

Acknowledgements

We are grateful to A. Bankier, C. Brown, and C. Fuller for helpful discussions and advice.

References

1. Murphy, G. and Ward, E. S. (1989). In *Nucleic acids sequencing: A practical approach* (ed. C. J. Howe and E. S. Ward), pp. 99–115. IRL Press at Oxford University Press, Oxford.
2. Mitchell, L. and Merril, C. R. (1989). *Analytical Biochemistry*, **178**, 239.
3. Innis, M. A., Myambo, K. B., Gelfand, D. H., and Brow, M. D. (1988). *Proceedings of the National Academy of Sciences, USA*, **85**, 9436.
4. Messing, J. and Bankier, A. T. (1989). In *Nucleic acids sequencing: A practical approach* (ed. C. J. Howe and E. S. Ward), pp. 1–36. IRL Press at Oxford University Press, Oxford.
5. Bankier, A. T. and Barrell, B. G. (1989). In *Nucleic acids sequencing: A practical approach* (ed. C. J. Howe and E. S. Ward), pp. 37–78. IRL Press at Oxford University Press, Oxford.
6. Heiner, C. and Hunkapiller, T. (1989). In *Nucleic acids sequencing: A practical approach* (ed. C. J. Howe and E. S. Ward), pp. 221–35. IRL Press at Oxford University Press, Oxford.
7. Gussow, D. and Clackson, T. (1989). *Nucleic Acids Research*, **17**, 4000.
8. Bankier, A. T. and Barrell, B. G. (1983). In *Techniques in the life sciences*, Vol. B5 (ed. R. A. Flavell), pp. 1–35. Elsevier Scientific Publishers, Ireland.

9. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). *Proceedings of the National Academy of Sciences, USA*, **80**, 3963.
10. Ward, E. S. and Howe, C. J. (1989). In *Nucleic acids sequencing: A practical approach* (ed. C. J. Howe and E. S. Ward), pp. 79–97. IRL Press at Oxford University Press, Oxford.
11. Bishop, M. J. (1989). In *Nucleic acids sequencing: A practical approach* (ed. C. J. Howe and E. S. Ward), pp. 185–219. IRL Press at Oxford University Press, Oxford.
12. von Heijne, G. (1987). *Sequence analysis in molecular biology*. Academic Press, San Diego.