

Demystified ...

DNA nucleotide sequencing

J S Randhawa, A J Easton

The determination of the nucleotide sequence of a gene is often the ultimate goal of a research project and the frequency with which new sequences are described in the literature and computer databases is remarkable. The basic techniques that underpin these endeavours are remarkably simple and their description now frequently forms part of the very earliest portion of a biological science undergraduate degree. Despite this, many misunderstandings about the nature of the reactions occur. In this review, we consider the mechanics of the techniques, including aspects of the associated technologies such as polyacrylamide gel electrophoresis. We describe the type of DNA suitable for sequencing and how it may be prepared, the details of the enzymatic reactions that can be used by the researcher, and the modifications to the techniques that have allowed the development of automation, which has made the goal of determining the nucleotide sequence of the entire human genome a reality.

The determination of DNA nucleotide sequences

The ultimate description of a gene is the determination of the sequence of nucleotide bases of which it is comprised and which, in turn, determines the nature of the encoded protein product. From the first development of an accessible and routine technique to determine the nucleotide sequences of only small stretches of DNA we have now advanced to the stage where nucleotide sequences are determined at a prodigious rate. This has resulted in the sequencing of the complete genomes of a number of bacteria and the eukaryote yeast *Saccharomyces cerevisiae*. Such is the pace of progress in the application of nucleotide sequencing technology that the sequence of the complete human genome will be available within the next five to seven years. This remarkable progress has been made possible by the application of relatively simple technologies. Such was the importance of these techniques that their proponents, Walter Gilbert and Fred Sanger, were jointly awarded the Nobel Prize for chemistry in 1980. The determination of the nucleotide sequence of DNA has become sufficiently routine that research papers describing new sequences usually do not contain full descriptions of the methods used. Over recent years, much emphasis has been placed on the automation of DNA sequencing techniques. The result of this is that

many researchers who use the techniques are not familiar with their basis. In this review, we describe the principles of the sequencing techniques and consider their advantages and drawbacks. We also describe the ancillary techniques, such as polyacrylamide gel electrophoresis, which underpin all of the analyses. In doing this we hope to make clear the fundamental aspects of the techniques so that one can decide on the most appropriate approach.

The chemical degradation method of DNA sequencing

The earliest method developed for determining the sequence of considerable lengths of either single stranded or double stranded DNA (ssDNA and dsDNA, respectively) was that of chemical degradation of a predefined length of DNA.¹ For this protocol a DNA fragment was isolated, usually from a recombinant plasmid. The fragment was then radiolabelled, at one end only, with a ³²P phosphate group, either at the 5' end using bacteriophage T4 polynucleotide kinase, or at the 3' end, by incorporation of a radiolabelled nucleotide (fig 1). This allows the subsequent detection of the reaction products. The Maxam and Gilbert DNA sequencing method requires the duplex DNA to be linear and only one strand to be labelled with ³²P. A series of chemical reactions are carried out, each designed to generate a cleavage after a specific base or pair of bases. The first step is to ensure modification of the base(s) in question so that the phosphodiester bond following them becomes susceptible to cleavage; for example, methylation of a guanine residue by dimethyl sulphate (DMS). Other bases can be altered individually or in combination with the use of chemicals such as hydrazine (which modifies both cytosine and thymidine bases), NaOH (which modifies both adenine and cytosine, although it has a preference for the former), piperidine formate (for adenine and guanine bases), and 1.5 M NaCl and hydrazine (for cytosine). By conducting several reactions in tandem it is possible to determine which base in a pair has been modified at a particular position in the DNA. Reaction conditions are controlled such that partial modification of the DNA bases occurs, the aim being to modify each DNA molecule in the reaction at least once. Once the modification reactions have been carried out, strand specific cleavage can take place in the presence of piperidine. This

Department of
Biological Sciences,
University of Warwick,
Coventry
CV4 7AL, UK
J S Randhawa
A J Easton

Correspondence to: Dr
Easton.

Accepted for publication
22 September 1998

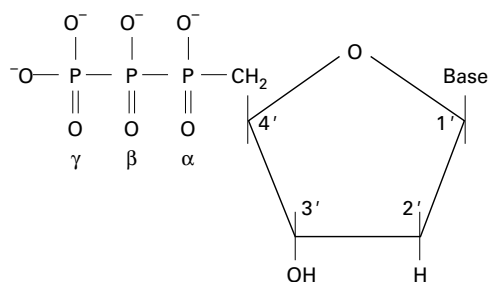


Figure 1 Structure of a nucleotide triphosphate. The attached base can be adenine, guanine, thymine, or cytosine. In dideoxynucleotides the 3' hydroxyl group (OH) is replaced by a H group. For radioactive labelling of the 5' end of DNA the isotope ^{32}P is located at the γ position. For labelling DNA at the 3' end the radioisotope is located at the α position.

produces a set of products for each reaction in which the length of each product is determined by the distance of the cleavage site, and hence the target nucleotide, from the radioactively labelled terminus. Following denaturation to single strands, the products of four to five individual reactions are separated on the basis of their size by electrophoresis on a polyacrylamide gel (see below). The products are then detected by autoradiography. An example of such a series of reactions is shown in fig 2. The nucleotide sequence of the DNA fragment can then be determined by reading from the bottom of the gel, where the smallest products produced by cleavage near the site of labelling migrate towards the top. Comparison of appropriate pairs of reactions unambiguously identifies the sequence.

Although this approach opened up the possibility of determining the sequences of genes there were some practical difficulties. The main drawback was that the control of the modification reactions was critical: over-reaction resulted in very small products, which limited the length of the sequence determined, and under-reaction with incomplete modification generated only large products which, by virtue of the linear gel systems in use at the time, were very difficult if not impossible to resolve. The precise timing of the reactions was affected by the length of the DNA fragment and considerable experience was required to obtain a good sequence. On average, each experiment determined the sequence of 150–200 bases. Because the point at which the DNA was labelled was determined by the location of a fortuitous restriction endonuclease cleavage site, elaborate protocols for the preparation of the target DNA were often necessary. This made the process very arduous and slow. In addition, the toxic nature of some of the chemicals used in the reactions was undesirable. For these reasons, this protocol is no longer the first choice for molecular biologists and has been superseded by the easier chain terminator method of sequencing. However, the chemical degradation method continues to be used in specific circumstances.

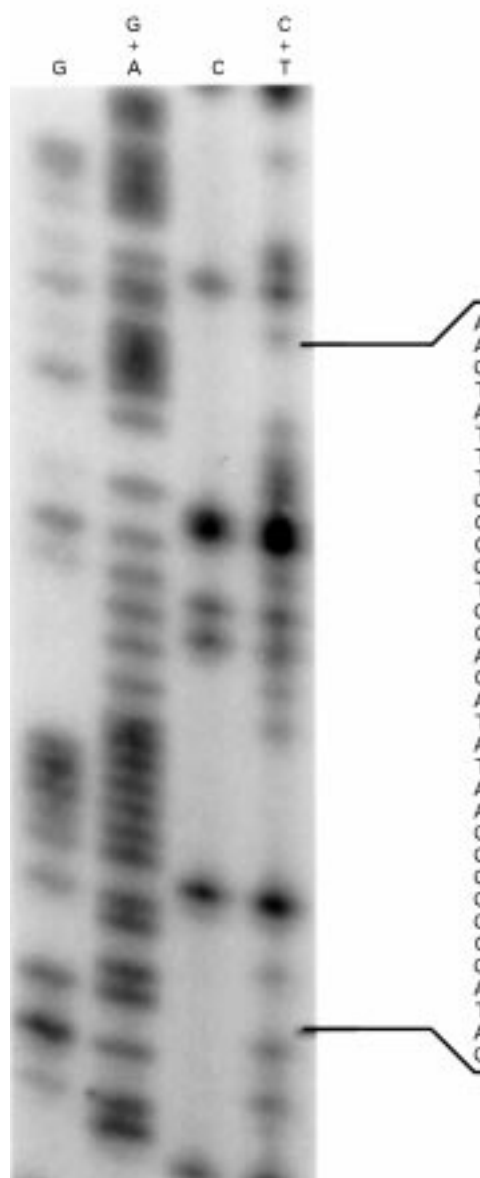


Figure 2 Example of a DNA sequence determined by the chemical degradation method of Maxam and Gilbert.¹ A DNA fragment radioactively labelled at one end has been subjected to four separate reactions that modify either guanosine alone (G), guanosine and adenine (G + A), cytosine alone (C), or cytosine and thymine (C + T). The products have been separated by polyacrylamide gel electrophoresis. The sequence is indicated on the right and is determined by comparison of the presence of reaction products in pairs of reactions: G with G + A, C with C + T. The sequence is read from the bottom of the gel to the top.

The chain terminator method for DNA sequencing

This method was first developed by Sanger *et al* and has become the basis for sequencing of DNA in laboratories worldwide.² The method uses an enzymatic reaction in which a chemically synthesised oligonucleotide primer is annealed to either ssDNA, or to a specific strand of duplex DNA that has been denatured, to form a short region of dsDNA. Typi-

cally, the primer is 17–25 nucleotides in length. The method relies on the ability of DNA polymerases to use the denatured strand of DNA as a template and synthesise faithfully the complementary strand by chain elongation of the primer in the presence of deoxynucleoside triphosphates (dNTPs) *in vitro*. These polymerases have the ability to incorporate, into the growing chain, nucleoside analogues such as 2',3' dideoxynucleotide triphosphates (ddNTPs), which lack the 3' OH residue of the deoxyribose backbone (fig 3). The incorporation of these nucleoside analogues results in the termination of chain elongation by preventing phosphodiester bond formation with the next dNTP. Thus, in the presence of dNTPs and a small concentration of one of the ddNTPs, the primer is extended until the analogue is incorporated within the growing chain. By mixing the analogue with the authentic nucleoside in an appropriate proportion, the incorporation of the analogue can be controlled to ensure that it is inserted only occasionally in the growing chain, with the frequency of insertion being determined by the ratio of the two related nucleosides. On average, if the authentic nucleoside predominates, the chain will extend a considerable distance before a terminating analogue is inserted whereas, if the analogue predominates, the growing chains will terminate very quickly after extension has begun. If conditions are correct, the result will be the accumulation of a randomly terminated, nested set of DNA molecules (fig 4). The products of the reaction are labelled for detection, either by using a labelled primer or by incorporation of a labelled nucleoside. This is discussed below. Four separate reactions in which the chain has been terminated with either ddGTP, ddATP, ddTTP, or ddCTP are performed. The double stranded products are denatured by heating and separated by electrophoresis, as for the chemical degradation method. The nucleotide sequence can then be read easily from the gel because each fragment must be terminated by the known analogue present in each reaction. A representative result is shown in fig 5.

As with all procedures, the original technique had some drawbacks. Regions of high DNA secondary structure may not remain completely denatured during electrophoresis. This results in a portion of the gel that contains several products running to the same position, and this shows up as regions of compression, or "four tracking", in which the precise sequence cannot be determined. To overcome this problem, 7-deaza-2'-deoxyguanosine-triphosphate (deaza-dGTP) or 2'-deoxyinosine-5'-triphosphate (dITP) can be substituted in the sequence reaction. These compounds form weak base pairings and DNA containing them is therefore easily denatured. If the problem still occurs, the sequence can only be determined from the opposite strand of DNA. It is standard practice to always determine the sequence of the segment of DNA from both strands. Although the ratio of each analogue to its cognate authentic nucleoside is important and must be determined empirically, in practice this is not difficult. By altering the ratios it

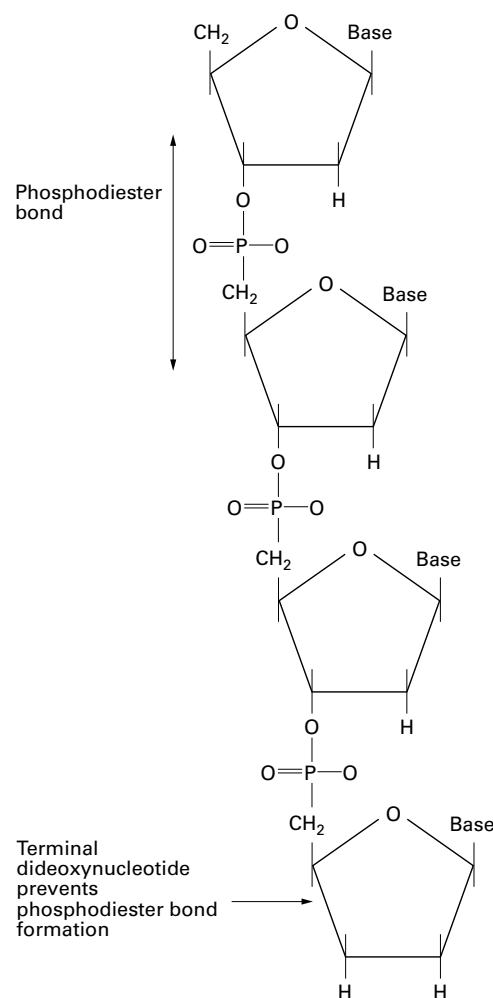


Figure 3 Diagrammatic representation of the structure of a segment of single stranded DNA. The phosphodiester bond is indicated. The insertion of a 2',3' dideoxynucleotide triphosphate into the DNA prevents phosphodiester bond formation.

is possible to bias the point of termination to one that is near or far from the primer site. This means that it is possible to determine the sequence over much longer regions of DNA than is possible with the degradation technique. In a simple electrophoresis step it is possible to determine an average of 250 bases, with the limiting feature being the gel system. Such is the simplicity of the principle of the chain termination technique that it forms the basis for all protocols now in use.

DNA polymerases

Initially, DNA sequence reactions were performed using the Klenow fragment of *Escherichia coli* DNA polymerase I. The Klenow fragment maintains the 5'→3' polymerase and 3'→5' exonuclease activities, but not the 5'→3' exonuclease activity, of the intact enzyme.³ Protocols using this enzyme can be found in most molecular biology manuals such as Sambrook *et al.*⁴ The Klenow fragment is not suitable for sequencing more than ~250 bases because of its low processivity and high tendency to prematurely disassociate from the DNA template. This results in the accumulation of prematurely terminated fragments of

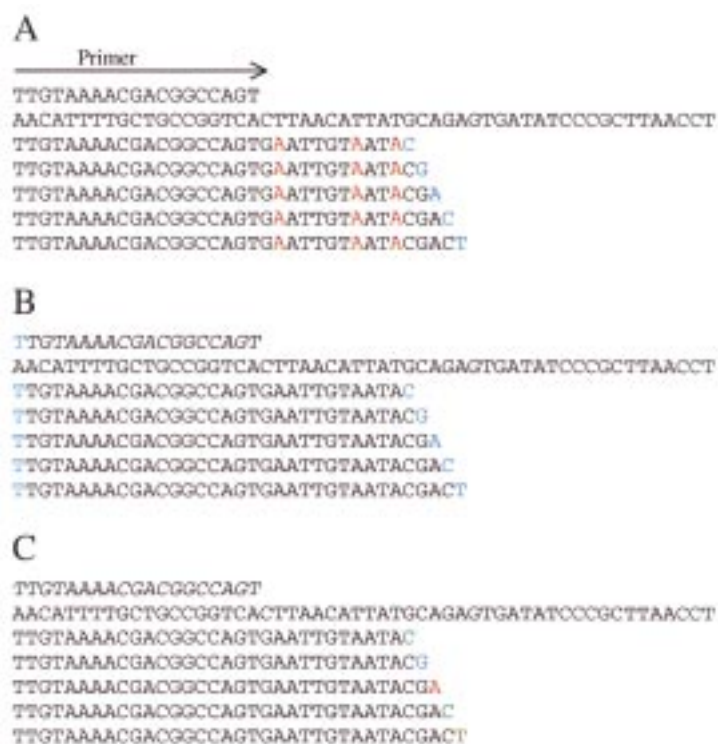


Figure 4 Example of the termination of chain elongation method on a known sequence by incorporation of each of the four ddNTPs, where: (A) the radioisotope is incorporated within the extended chain; (B) the 5' end of the extended chain is labelled radioactively with ^{32}P ; and (C) each terminating ddNTP is labelled with a unique fluorescent label.

DNA not terminating with the nucleoside analogue, which show up as high background on an autoradiograph. Furthermore, the Klenow fragment encounters difficulty when reading homopolymeric sequences or regions containing strong DNA secondary structure.

Alternatively, the T7 DNA polymerase can be used for manual sequencing because it has high processivity. More than 500 bases can be read from sequencing reactions carried out with this enzyme, although polyacrylamide gel limitations only allow up to 400 bases to be read satisfactorily (see below). A genetically engineered form of the T7 polymerase (Sequenase™ version 2.0⁵ 6) is used in the highly popular Sequenase™ sequencing kit available from Amersham plc (Amersham, Buckinghamshire, UK). The 3'→5' exonuclease activity of the enzyme has been completely removed. Sequenase and native T7 DNA polymerase are also more tolerant of deaza-dGTP and dITP substitutions in the sequence reaction. Other polymerases used in manual sequencing are the reverse transcriptases and those isolated from thermophilic microorganisms, such as Taq DNA polymerase from *Thermus aquaticus* YT1; Vent_R (exo⁻), a modified form of the DNA polymerase from *Thermococcus litoralis* (New England Biolabs, Beverly, Massachusetts, USA); and Thermo Sequenase DNA polymerase⁷ (Amersham plc). The latter types of polymerases are used in cycle sequencing reactions (see below).

DNA template preparation

The choice of template preparation is to some extent determined by the sequencing strategy

developed by the researcher. Initially, the choice is whether to sequence ssDNA or duplex DNA.

Invariably, ssDNA is obtained using bacteriophage M13 based cloning vectors. Bacteriophage M13 was first used as a vector in the 1970s⁸ and subsequently developed into the commonly used M13 “mp” series.⁸⁻¹² Only F' or Hfr strains of *E coli* are infected by filamentous M13 bacteriophage, and virus particles contain only a single strand of circular DNA. Upon infection, the circular ssDNA is converted to circular dsDNA (replicative form; RF). From this RF DNA, ssDNA homologous to only one of the strands is produced and is packaged into infectious phage particles. The duplex DNA of M13 can be used to clone foreign DNA sequences and, when transformed into *E coli*, becomes the RF form of M13, yielding phage carrying the foreign DNA sequence.

Isolation of single stranded M13 DNA is a simple matter of infecting a culture of *E coli* overnight and precipitating phage particles from the supernatant in the presence of polyethylene glycol 600 and NaCl. Purified phage particles are then extracted with phenol and chloroform before precipitation of the DNA. No chemical denaturation of M13 DNA is required because the DNA is already single stranded. However, in general, heating of the ssDNA with the primer is required. Universally available primers have been designed specifically for sequencing using ssDNA as the template. ssDNA from M13 gives very good sequence data using either of the sequencing methods mentioned above. Although M13 ssDNA gives sequence data of the highest quality, large DNA fragments have frequently been found to be unstable when introduced into the phage. Sequencing of M13 ssDNA is also unidirectional and adds a time consuming subcloning step to the process.

Sequencing of inserts cloned into plasmids can permit the analysis of both strands of DNA. Plasmids are also more tolerant of larger inserts and represent a direct way of determining the sequence of inserts that have been cloned for other purposes. However, plasmid DNA needs to be of high quality when used for DNA sequencing. Simple alkaline lysis preparation¹³ of plasmid DNA may yield DNA of sufficient quality for sequencing. The boiling lysis method of plasmid DNA preparation¹⁴ cannot be applied to all strains of *E coli* because the protocol does not remove endonuclease from EndA strains. DNA purified from caesium chloride gradients is of the highest quality. A more recent technique of plasmid isolation is the use of silica, which can bind DNA reversibly under certain conditions and can be used for isolation of DNA. Binding to silica forms the basis of many commercially available plasmid DNA isolation kits. Automated sequence reactions (see below) are more sensitive to contaminants introduced during the purification of plasmid DNA and traces of salt, ethanol, and phenol drastically reduce the quality of the sequence obtained from automated machines. Manual methods of DNA sequencing appear to be more robust in their tolerance of these contaminants. DNA prepared using the

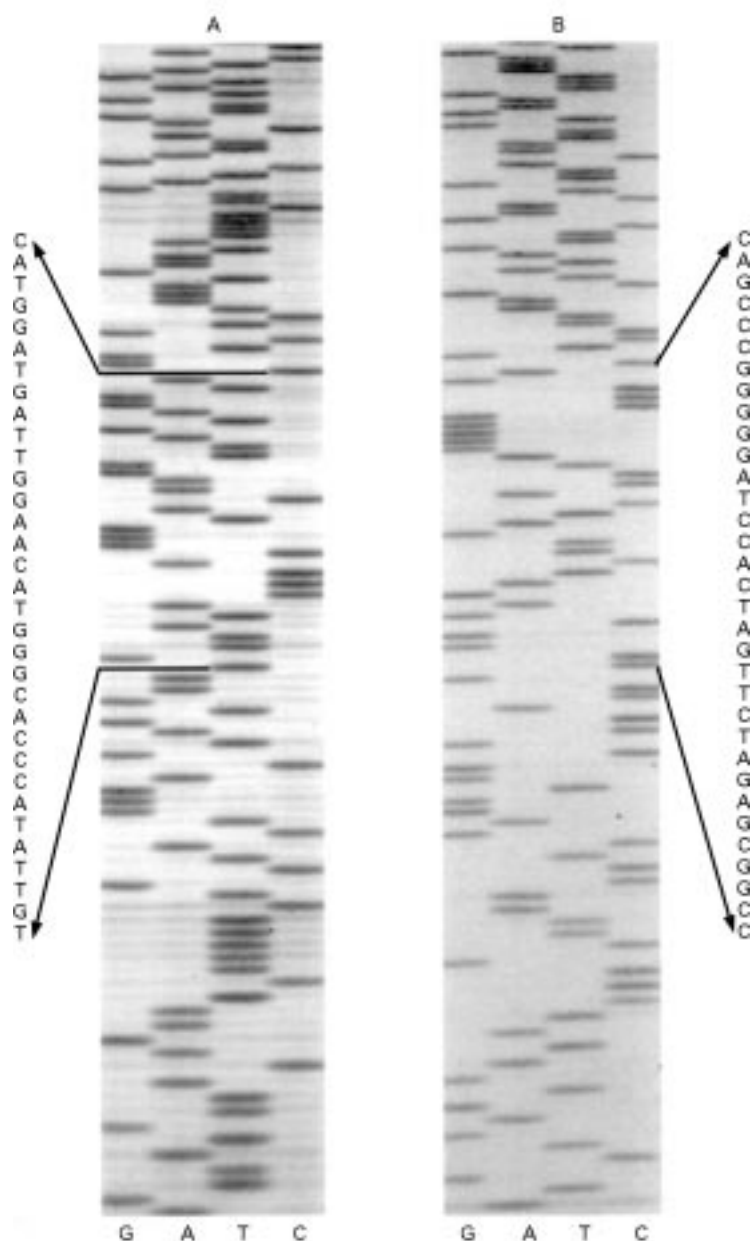


Figure 5 Autoradiograph of a DNA sequence gel obtained by the chain termination method. The chain terminating nucleotides in each of the four separate reactions are indicated. The nucleotide sequence is read from the bottom to the top. (A) An example of sequence data obtained using Sequenase version 2.0; and (B) sequence obtained after cycling in a thermal cycler (Thermo Sequenase).

Qiagen plasmid preparation kits (Qiagen GmbH, Germany) or the Wizard[®] Plus (Promega, Wisconsin, USA) kits yield DNA of high quality that is suitable for sequencing.

For sequencing of polymerase chain reaction (PCR) products, it is recommended that the user purifies the fragment of interest from an agarose gel. The excised band can be purified by the use of commercially available kits such as the Qiagen gel extraction kit or the Promega Wizard DNA clean up kit. Alternatively, DNA can be recovered by simply centrifuging the excised band through an Eppendorf tube with a small hole located at its base. The DNA can then be ethanol precipitated and used in sequence reactions.

dsDNA template denaturation

Both strands of duplex DNA need to be separated before primer binding and sequence reaction. Several protocols have been developed to ensure the complete denaturation of duplex DNA. The most commonly used protocols involve the incubation of dsDNA for 30 minutes at 37°C with 0.1 volumes of 2 M NaOH and 2 mM Na₂EDTA to ensure complete denaturation. The next step requires the neutralisation of alkali to ensure the proper sequence reaction conditions are achieved. Neutralisation can be achieved by the addition of 0.1 volumes of 3 M sodium acetate (pH 4.5–5.5), followed by ethanol precipitation.¹⁵ The denatured DNA is then annealed to the primer and the sequence reactions performed. Alternatively, denaturation can be performed with 1 µl of 1 M NaOH and neutralised by the addition of 1 µl 1 M HCl.¹⁶ This overcomes the tedious step of ethanol precipitation. However, care must be taken to ensure that the correct amount of HCl is added so that the sequence reactions are not performed under acidic conditions. A quick but effective way is to mix 10 µl of 1 M NaOH with 10 µl of 1 M HCl. Sequence reaction buffer is added to give a final concentration of ×1. If the pH of this mix is the same as the normal reaction mixture then the concentrations of NaOH and HCl can be considered to be balanced. If not, then it is just a matter of adjusting the amount of HCl to NaOH and re-assaying: pH sticks can be very useful for this approach. Another protocol for neutralisation involves centrifuging the denatured DNA (in NaOH) through a dialysis spin column.

An alternative to chemical denaturation of dsDNA is the use of high temperature. Several commercially available DNA sequencing kits known as cycle sequencing kits have combined the use of thermophilic DNA polymerases and heat denaturation. These kits take advantage of the ability of DNA polymerases, isolated from thermophilic bacteria, to carry out chain elongation at high temperatures. Typically, DNA and primer are mixed with a DNA polymerase, reaction buffer, radiolabelled nucleoside triphosphate, and a balanced mixture of dNTPs and ddNTPs. The reactions are subjected to 30–40 cycles of heating and cooling. One cycle consists of template denaturation by heating to 95–100°C, followed by primer annealing at 50–60°C, and primer extension at the optimal temperature for polymerase activity. Because of the high temperatures involved in these protocols, primers have to be designed with relatively high melting temperatures compared with those used in alkali denaturation protocols. In addition, cycle sequencing kits require the use of a thermal cycler, which can add to the cost of the sequencing project. However, most laboratories that carry out DNA sequencing also have thermal cyclers. One cautionary note is the choice of radioisotope used in these reactions. The advantages and disadvantages of using either ³²P or ³⁵S are discussed below. Cycle sequencing with ³⁵S radiolabelled nucleotides can result in the contamination of the thermal cycler owing to the presence of volatile

radioactive impurities. This problem does not appear to manifest itself when ^{32}P or ^{33}P are used for labelling nucleotides.

Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis, to separate the reaction products of sequencing reactions, underpins all of the techniques described above. The DNA fragments are resolved according to their sizes and gels containing 6% acrylamide are generally appropriate for the separation of DNA fragments between 25 and 500 bases long.

Each sequence reaction is heat denatured at 80–100°C for two to three minutes before electrophoresis to separate the DNA strands and to denature any localised areas of secondary structure. During electrophoresis, heat is generated as the running buffer passes through the polyacrylamide gel matrix. In the presence of urea, sufficient electrical current is supplied to keep the temperature at 60–70°C. This combination of heat and urea keeps the DNA molecules denatured during electrophoresis. For this reason, before loading of the sequence reactions, polyacrylamide gels are run until they achieve the correct temperature. This ensures that there is only a small period of time when the denatured DNA can reanneal and secondary structures re-form. In addition, the procedure removes any ionic products formed during polymerisation. A consequence of running gels at such a high temperature is uneven dissipation of heat by the glass plates, which results in the samples at the outer edges migrating more slowly than those located in the middle of the gel, giving the commonly observed “smiling” effect. This can be overcome by dissipating heat evenly across the whole of the gel by placing one of the glass plates containing the gel in contact with a metal plate. Sequencing gel apparatus is available where the back glass plate is in contact with a sheet of aluminium. Another, less common, way of overcoming this problem is to use a water jacket placed against one of the glass plates.

Detection of nested sets of DNA fragments

Although the chemistry behind the chain termination method remains the same, the method of detection of the nested DNA molecules has altered. The most commonly used method for detection of the DNA molecules is by the incorporation of a radiolabel into each molecule. The radiolabel can be attached either to the 5' end of the primer by a simple kinase reaction before its use in the extension reaction, or by incorporating radiolabelled dNTP into the growing chain itself (fig 4). This is a simple matter of spiking the reaction mixture with a radiolabelled nucleoside triphosphate. One should be aware that incorporating radiolabel into the growing strand results in smaller fragments showing up as fainter bands, whereas larger DNA fragments appear as stronger bands further up the gel. When the primer itself is radiolabelled, all the bands should appear to be of even intensity. In

most cases, the nucleoside is labelled with ^{32}P or ^{35}S (fig 1). The choice of isotope depends upon the researcher's preference, although each isotope has advantages and disadvantages.

^{32}P is a strong β particle emitter and has a half life of 14.3 days. This decreases the amount of time that gels need to be exposed to x ray film. Bands on an x ray film exposed to gels containing ^{32}P are frequently found to be large and diffuse, as a result of scattering of the strong β particles. This is a problem when reading near the top of a gel, where fragments run close together, or when differentiating between closely spaced bands. Also, strong β particle emission results in the radiolysis of DNA when stored for longer than one to two days, leading to degradation of the DNA and unreadable gels. The use of ^{35}S labelled dNTP can overcome these problems. ^{35}S is a weak β particle emitter with a half life of 87.4 days. The lower energy β particles do not cause radiolysis to the same extent, and are not scattered to the same extent, as those of ^{32}P , giving sharper, well defined bands. Thus, ^{35}S labelled sequence reactions can be stored for longer before electrophoresis. One minor drawback of using ^{35}S as the isotope is the need to expose x ray films to the gel for longer times than are necessary for ^{32}P . More recently, an alternative has been the use of nucleosides labelled with ^{33}P . DNA fragments incorporating ^{33}P have similar characteristics to those labelled with ^{35}S . However, ^{33}P has a half life of 25.4 days and gives darker bands after autoradiography than does ^{35}S labelled DNA. Furthermore, ^{33}P labelled dNTPs are incorporated much more efficiently within the growing DNA strand than are ^{35}S labelled nucleosides.

One of the drawbacks of using radiolabelled dNTPs in DNA sequencing is the requirement of special procedures for the disposal of radiolabelled material. The use of silver staining for visualisation of DNA fragments separated on an acrylamide gel dispenses with the use of radioisotopes. However, this procedure requires the entire acrylamide gel to be soaked in a solution of silver nitrate and formaldehyde and then fixed in an alkaline solution of sodium carbonate containing formaldehyde and sodium thiosulphate¹⁷; in the presence of formaldehyde silver ions are reduced to metallic silver. This reaction is stopped by the addition of acetic acid. DNA bands are then visualised on a white light box. A permanent record of the gel may require specialist equipment. This is not only time consuming, but also messy, expensive, and uses hazardous chemicals such as silver nitrate and formaldehyde. The Silver Sequence™ (Promega) kit is based upon this method.

Another alternative is the use of chemiluminescent methods. These procedures require the use of biotinylated DNA primers, with sequencing reactions and electrophoresis carried out as normal. Biotin labelled phosphoramidites are used during primer synthesis to add the biotin label. After electrophoresis, the biotinylated DNA fragments are transferred to a nylon membrane, UV crosslinked, and bound to a streptavidin-alkaline phosphatase conju-

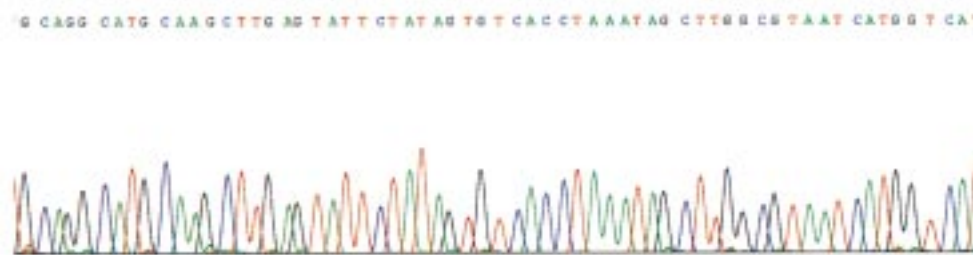


Figure 6 Computer generated output from an automated DNA sequencer. The data show the intensity of each fluorescent band detected underneath the corresponding base. Sequence data can then be imported into a software analysis package.

gate. After washing off excess conjugate, DNA is incubated with a chemiluminescent substrate of alkaline phosphatase. Nylon membranes are then exposed to x ray film and the DNA fragments detected.¹⁸ Alternatively, DNA probes linked to biotin or alkaline phosphatase can be hybridised to the DNA bands before chemiluminescent detection. Although this procedure might appear to be simple and effective, extreme care is required in the handling of the nylon membrane and in the transfer of DNA fragments from the polyacrylamide gel matrix to the membrane.

Typically, from setting up of the sequencing reaction to running the gel can take up to two to three days. Depending upon the incorporation of the radiolabel within the DNA molecules, exposure time can vary from one to three days. A further day is then required for developing and analysing the autoradiograph. Once the sequence has been determined, reactions might need to be repeated to obtain further sequences or just re-run for longer on an acrylamide gel. Further time is required to transfer the sequence data obtained into a format that can be used for analysis by computer software. From these gels, up to 400 bases of readable and good quality data can be obtained. For the sequencing of large inserts it may be necessary to subclone fragments of the DNA target into plasmid vectors or to design primers complementary to the sequence data obtained from previous analyses.

Automated DNA sequencing

Recent developments have focused on making the sequencing reactions and characterisation of products easier. There are now commercially available, and widely used, automated sequencing machines that combine the initial running of samples and the reading of tracks to generate the nucleotide sequence. These machines are designed for high throughput of DNA sequencing and require the use of a unique set of nucleoside analogues in conjunction with the chain termination reaction. The principle of the reaction is as described above. Attached to each dideoxy nucleoside triphosphate is a fluorescent dye that, when excited by illumination with light of an appropriate wavelength, emits light at a different wavelength. The automated sequence reaction takes place in a single tube because each nucleoside analogue is attached to a different fluorescent dye. The reaction uses thermophilic DNA polymerases in the cycle sequencing reaction.

When the products of the reaction are separated by electrophoresis, each DNA molecule migrates, according to its size, towards the bottom of the gel. Light from a laser is used to excite the fluorescent dye as the DNA fragment bearing it migrates past. The fluorescent light emitted from each dye is detected by a photomultiplier and converted into an electrical signal, which is then interpreted by computer software as the appropriate nucleotide base. The strength of fluorescent light determines the strength of the electrical signal. The chemistry in this process is slightly different because only DNA molecules terminated with a fluorescent dye analogue are detected. Prematurely terminated DNA molecules are not detected by this system because no dye analogues are incorporated into the chain: this reduces background. This system also enables the user to collate quickly the nucleotide sequence data, which is stored in electronic format, and continue with the analysis. The computer system not only expresses the data in terms of the fluorescent signal detected, but also as the DNA sequence, and can display both together, as shown in fig 6. The length of sequence determined can be as much as 700–1000 bases in each run. The main disadvantage of this system is the very high cost of equipment required for the running, detection, and analysis of DNA fragments. For this reason, automated sequencing is normally available as a core facility within a research institute, university, or company. The higher price for sequence reactions offsets the time saved in traditional methods of DNA sequence analysis.

In most cases, it is the amount of resources available to each individual researcher that determines the experimental approach undertaken to determine the DNA sequence. It is highly recommended that the researcher uses one of the commercially available DNA sequencing kits to begin with. Once familiar with the basic techniques, DNA sequencing costs can be reduced by the use of solutions and mixtures prepared in house. Readers should be aware that although DNA sequencing is performed routinely in most laboratories, protocols exist for the direct sequencing of RNA. However these are outside the remit of this review.

1 Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977;74:560–4.

- 2 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463-7.
- 3 Klenow H, Henningsen I. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* β by limited proteolysis. *Proc Natl Acad Sci USA* 1970;65:168-75.
- 4 Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989.
- 5 Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* 1987;84:4767-71.
- 6 Tabor S, Richardson CC. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *J Biol Chem* 1989;264:6447-58.
- 7 Tabor S, Richardson CC. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxynucleotides. *Proc Natl Acad Sci USA* 1995;92:6339-43.
- 8 Messing J, Gronenborn B, Muller-Hill B, et al. Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind II fragment of the lac regulatory region in M13 replicative form in vitro. *Proc Natl Acad Sci USA* 1977;74:3642-6.
- 9 Messing J, Crea R, Seeburg PH. A system for shotgun DNA sequencing. *Nucleic Acids Res* 1981;9:309-21.
- 10 Gronenborn B, Messing J. Methylation of single-stranded DNA in vitro introduces new restriction endonuclease cleavage sites. *Nature* 1978;272:375-7.
- 11 Messing J, Vieira J. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 1982;19:269-76.
- 12 Norrander J, Kempe T, Messing J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 1983;27:101-6.
- 13 Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 1979;7:1513-23.
- 14 Holmes DS, Quigley M. A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* 1981;114:193-7.
- 15 Lim HM, Pene JJ. Optimal conditions for supercoil DNA sequencing with the *Escherichia coli* DNA polymerase I large fragment. *Gene Analysis Techniques* 1988;5:32-9.
- 16 Hsiao K. A fast and simple procedure for sequencing double stranded DNA with Sequenase. *Nucleic Acids Res* 1991;19:2787.
- 17 Bassam BJ, Caetano-Anolles G, Gresshoff PM. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 1991;196:80-3.
- 18 Martin C, Bresnick L, Juo RR, et al. Improved chemiluminescent DNA sequencing. *BioTechniques* 1991;11:110-13.

Journal of Clinical Pathology - <http://www.molpath.com>

Visitors to the world wide web can now access the *Journal of Clinical Pathology* either through the BMJ Publishing Group's home page (<http://www.bmj.com>) or directly by using its individual URL (<http://www.molpath.com>). There they will find the following:

- Current contents list for the journal
- Contents lists of previous issues
- Members of the editorial board
- Information for subscribers
- Instructions for authors
- Details of reprint services.

A hotlink gives access to:

- BMJ Publishing Group home page
- British Medical Association web site
- Online books catalogue
- BMJ Publishing Group books.

The web site is at a preliminary stage and there are plans to develop it into a more sophisticated site. Suggestions from visitors about features they would like to see are welcomed. They can be left via the opening page of the BMJ Publishing Group site or, alternatively, via the journal page, through "about this site".



Demystified ... DNA nucleotide sequencing.

J S Randhawa and A J Easton

Mol Path 1999 52: 117-124

doi: 10.1136/mp.52.3.117

Updated information and services can be found at:

<http://mp.bmj.com/content/52/3/117.citation>

References

These include:

Article cited in:

<http://mp.bmj.com/content/52/3/117.citation#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>