Enzymatic multiplex DNA sequencing

Mark Chee
Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received March 15, 1991; Revised and Accepted May 2, 1991

ABSTRACT

The problem of reading DNA sequence films has been reformulated using an easily implemented, multiplex version of enzymatic DNA sequencing. By utilizing a uniquely tagged primer for each base-specific sequencing reaction, the four reactions can be pooled and electrophoresed in a single lane. This approach has been previously proposed for use with fluorescently labelled probes (1), and is analogous to the principle used in four-dye fluorescence sequencing except that the signals are resolved following electrophoresis (2). After transfer to a nylon membrane, images are obtained separately for each of the four reactions by hybridization using oligonucleotide probes. The images can then be superimposed to reconstitute a complete sequence pattern. In this way the correction of gel distortion effects and accurate band registration are considerably simplified, as each of the four base-specific ladders require very similar corrections. The methods therefore provide the basis for a second generation of more accurate and reliable film reading programs, as well as being useful for conventional multiplex sequencing. Unlike the original multiplex protocol (3), the approach described is suitable for small projects, as multiple cloning vectors are not used. Although more than one vector can be utilized, only a library of fragments cloned into any single phage, phagemid or plasmid vector is actually required, together with a set of tagged oligonucleotide primers.

INTRODUCTION

The community of biologists is undertaking the sequencing of representative genomes of various free-living organisms, ranging in size from Mycoplasma (800kb) to mammals (3 Gb) (4). However, the largest contiguous DNA sequences which have been determined so far are the genomes of several dsDNA eukaryotic viruses (5, 6, 7, 8, 9) and plant chloroplasts (10, 11, 12). The largest of these is the 229kb genome of human cytomegalovirus (8). The difficulty in sequencing millions of base pairs of DNA is that several steps in the methods are relatively labour intensive, although the sequencing reactions themselves are rapid and easily performed. Two limiting steps in conventional procedures are the size fractionation of sequencing reaction products by gel electrophoresis and the subsequent reading of sequence ladders. The former problem can be overcome by multiplexing, which theoretically allows an enormous amount of data to be obtained from a single gel by processing clones as mixtures rather than individually (3). Each sequence in the mixture is labelled by a unique short oligonucleotide 'tag' sequence. This allows the mixture to be resolved following electrophoresis: the superimposed sequence ladders are blotted from the gel to a nylon membrane, and detected one at a time by hybridization using tag-specific oligonucleotide probes. In practice, at least 50 sets of sequences can be obtained from a single gel (3).

Unfortunately, a bottleneck in the multiplex procedure is the reading of sequence films. In previous large-scale sequencing projects this task has been performed with the aid of a sonic digitizer (13, 14). Although film reading programs have been under development for some time (15), and some programs are commercially available, their error rates are presently more variable and unpredictable than that of a skilled person and the accurate interpretation of film-imaged sequence ladders by computer programs is difficult to achieve in routine practice. Programs specifically designed to read multiplex films have an advantage. This is because a sequence image can be used as an 'internal standard' to help interpret other images derived from the same membrane (3). However, the original implementation of the multiplex strategy used chemical sequencing (16), which yields a more complex sequence ladder than the enzymatic dideoxynucleotide chain-termination method (17). Most successful large scale sequencing projects have used the chain-termination method and bacteriophage M13 vectors, which allows the routine production of clean and easily interpretable sequences (18). It was therefore decided to adapt the original multiplex protocol for use with enzymatic sequencing, using tagged primers.

MATERIALS AND METHODS

Eight oligonucleotide sequencing primers were synthesized, each 37 nucleotides in length. The 3′ end of each primer consists of the 17 nucleotide M13 universal priming sequence [GTAAAA-CGACGGCCAGT3′]. The 5′ ends of the primers bear different 20mer tag sequences (Figure 1). In four of the primers, UE01C, UP01C, UE02C and UP02C, these tags are complementary to the E01, P01, E02 and P02 probe sequences respectively (copied from the original 'plex' vectors (3)). A second set of four primers, UJOL14C, UJOL15C, UJOL16C and UJOL17C, have the following tag sequences: 5′ CAAGTTTGAGGATCTCATT, TATCAATATATATGGTTTGA, GTTGTGCTACCAAGGAA-GCA, and TGTCCTAGAGCTGTCACTT, respectively. The
oligonucleotides were gel-purified (19) and used to sequence ssDNA templates prepared by phenol extraction (20) or SDS denaturation (21). Conventional sequencing reactions were performed as previously described (20).

For hybridization experiments, radioactively labelled nucleotides were omitted from the sequencing reactions. Instead, the 2 μl of each nucleotide mix added to the reaction mixture consisted of the following: ‘A’ mix: 6.25 μM dATP, 62.5 μM ddATP; ‘C’ mix: 6.25 μM dCTP, 40 μM ddCTP; ‘G’ mix: 6.25 μM dGTP, 80 μM ddGTP; ‘T’ mix: 6.25 μM dTTP, 250 μM ddTTP; as well as 125 μM of each of the three other dNTPs in each mix. Apart from the use of these modified mixes, no changes were made to the conventional sequencing procedure (20).

Sequencing reactions were pooled and ethanol precipitated as appropriate. Precipitation in microtitre trays was carried out as follows: a mixture of 3.2 μl 3M NaAc pH 5.0 and 112 μl ethanol was dispensed to individual wells of a microtitre plate (Falcon 3911 or Corning 25855) using an 8-channel pipetor. Each set of four reactions was added to the EtOH/NaAc mixture, and the tray sealed using a Falcon 3073 plate sealer. The samples were mixed by inversion and stored at −20°C for 30 minutes. The DNA was collected by a 20 minute centrifugation at 4 000 rpm in an IEC Centra 3C centrifuge. The sealer was removed, and the plate inverted to discard the supernatant. After blotting the tray on tissue paper, 200 μl of 95% EtOH was added to each well.

The plate was covered with a plastic lid and centrifuged for 2 minutes. The EtOH was discarded and the plate inverted for several minutes on tissue paper, then left for 20 minutes to air dry. Precipitated samples were resuspended in 6 μl deionized water by vortexing on an SMI multi-tube vortexer for 1 minute. Samples were denatured and electrophoresed on 6% polyacrylamide buffer gradient gels as previously described (20).

Following electrophoresis, the gel was transferred to a dry piece of Whatman 3MM blotting paper, and placed on a second sheet of blotting paper supported on a glass plate and saturated in 4 × SSC (SSC: 150 mM NaCl, 15 mM trisodium citrate). This sheet was wicked in a tray containing 1 litre of 4 × SSC. The DNA was transferred to a nylon membrane (Amersham Hybond N) by capillary blotting overnight (22). DNA was fixed to the membranes by U.V. crosslinking (23).

Plex oligonucleotide probes were a kind gift of Dr. George Church. Probes were tailed at their 3′ ends using [α-32P]dCTP as previously described (3). For the preparation of digoxigenin (DIG) labelled probes, identical tailing reactions were carried out substituting 10pmols of DIG-11 dUTP (Boehringer Mannheim) for [α-32P]dCTP. Membranes were prehybridized for at least 10 minutes in 4 × SSC, 5 × Denhardts’ (0.1% (w/v) each of BSA (heated at 80°C for 30 minutes to inactivate any alkaline phosphatase activity), Ficoll (Pharmacia) and polyvinylpyrrolidone), 0.5% (w/v) SDS, 5 mM NaHPO4 (23). Hybridization was carried out in 25–50 μl of prehybridization buffer per cm² of membrane. The probe concentration was approximately 1 nM. After 1 h at 42°C, unbound probe was removed by five 1 minute washes at room temperature in 1 × SSC, 0.5% SDS (200 μl/cm² membrane). Radioactive blots were covered in Saran wrap and exposed to film immediately. Detection of DIG labelled probes used an anti-DIG antibody-alkaline phosphatase conjugate (Boehringer Mannheim) according to the manufacturer’s instructions, except that all volumes were reduced by 70% and the conjugate was used at a 1:10 000 dilution. Blots were developed in 25 μl of 100 mM Tris.Cl pH9.5, 100 mM NaCl, 50 mM MgCl2, 0.15 mM AMPPD ([3(2′-adenylmantine-4-methoxy-4(3′-phosphoryloxy)phenyl-1,2-dioxetane); Tropix/cm² for 30 minutes at 37°C, prior to exposure to film. Probes and dioxetane were stripped from the membranes by two 10 minute washes at 70°C with 0.2% SDS, 2 mM EDTA (200 μl/cm² membrane).

The hybridization and washing procedures were carried out in plastic bags. However, washing steps have also been performed with gentle agitation in a perspex tub (43 × 27 × 15 cm) mounted on a reciprocal shaker, with equivalent results. In the latter case a minimum wash volume of 500 ml was used. The use of a tub is more convenient for batch processing and should be straightforward to automate.

RESULTS

Autoradiograms revealed no difference in sequence quality when tagged primers were used instead of the 17mer universal primer in conventional [α-35S]dATP labelled sequencing reactions and in multiplex hybridization experiments using [α-32P]dCTP-tailed probes (results not shown). Experiments were then conducted to determine the feasibility of pooling the four base reactions for each clone and fractionating them in a single lane to obtain a superimposed but interpretable set of sequence ladders. The question addressed was whether or not difficulties in band registration might arise as a result of mobility differences between the different primer sequences and/or distortion of the membrane between probings. It is relevant that an automated film reader employing an internal standard requires that the nylon membrane does not undergo significant distortions between probings (George Church, personal communication). Clones were sequenced using the four tagged primers UE01C, UP01C, UE02C, and UP02C, one for each base reaction (Figure 1). The A, C, G and T reactions for each clone were pooled, and processed as described above. A complete set of sequence autoradiograms was obtained from four consecutive rounds of probing with [α-35S]dATP tagged primers. Alignment of the films showed that sequence-specific mobility effects and distortion of the membrane between probings were sufficiently minor to allow accurate registration of the bands, and hence accurate reading of the sequence. At least 200 nucleotides of sequence could be read accurately from a single clone by simply tracing the four sets of bands using different colours, overlaying the tracings, and reading the bands sequentially. In order to assess the practicality of reading the sequences by machine, the images were scanned to provide optical density profiles (Figure 2). These profiles were overlaid, and were found to be sufficiently in register to allow accurate interpretation of the sequence for at least 300 nucleotides. This was essentially the limit of resolution of the gel for accurate manual reading.

In order to ensure that the relatively minor mobility differences observed between the four primers were not coincidental to the oligonucleotides used, a second set of four tagged M13 universal primers was synthesised, this time incorporating 20 mer sequences derived from the genome of murine herpesvirus-68 (UJ0L14C, 15C, 16C, 17C). Sequencing reactions were performed using [α-35S]dATP to label the DNA directly. Various templates were sequenced, and in all cases correctly ordered sequence ladders were obtained following conventional electrophoresis in which the four reactions were run side-by-side (results not shown).

Initial hybridization experiments were conducted using [α-32P]dCTP tailed oligonucleotide probes. However, the use of
Figure 1. Approaches to enzymatic multiplex DNA sequencing. a) A set of sequence-tagged vectors can be used. The tag site is shown in red, and the insert to be sequenced in blue. However, the original plex vectors (3) are plasmids, and therefore amenable only to dsDNA sequencing. Sets of bacteriophage M13 vectors have been constructed bearing either one (32) or two [Chee, unpublished] of the plex tag sites flanking the polylinker, which can be used for this approach. b) The strategy used in this paper. In this case the tag site is carried on the primer. c) If tagged primers are used, there is no practical impediment to performing each base reaction using a different primer, as depicted. The reactions can then be pooled in any combination desired. The configuration shown, in which the four reactions are electrophoresed in a single lane, is designed to facilitate accurate band registration and reading by an automatic film reader. In order to read the sequence manually, base reactions would be run side-by-side. The logistics of processing the reactions are essentially the same with either configuration; the same number of probings are required.

Figure 2. Four overlaid one-dimensional optical density profiles for a single clone shown in two overlapping sections. The optical density profiles are unprocessed, except for a simple transform to correct for the relative displacement (translation and rotation) of the four images from which they are extracted. The profiles read 5' to 3' from right to left. Nucleotides positions 66 to 214 from the start of the universal priming site are shown. The sequence is that of Bluescribe M13+ (template DNA obtained by rescue with M13K07 helper phage (30)), and was determined using the primers UE01C, UP01C, UE02C, and UP02C for the T, C, G and A specific reactions respectively. Detection was by autoradiography following hybridization with [α-32P]dCTP tailed plex probes.
using sequence-labelled primers and fluorophor-labelled probes a proposal was recently put forward for multiplex sequencing (George Church, personal communication; 2, 27). For example, proposed for various forms of multiplex DNA sequencing tagged vectors (3), tagged primers have also been used or triggered background luminescence, and not by the level of signal. Although the original multiplex protocol was based on a set of successfully in an 8-plex system. The nonradioactive methods described have been used to be in the range of 20 to 50fmols per reaction. However, the obtain an easily interpretable sequencing ladder was estimated. Exposure times of 10 to 15 minutes were overcomes these disadvantages (24, 25, 26). We utilized a different bridging system with similar results. Digoxigenin (DIG) labelled oligonucleotide probes were detected using anti-DIG antibody-alkaline phosphatase conjugates and a chemiluminescent dioxetane substrate. Exposure times of 10 to 15 minutes were typically required, following a one hour preincubation period (Figure 3). In our hands the DIG bridging system was similar in sensitivity to the streptavidin based system (24), and the practical lower limit of template ssDNA required in order to obtain an easily interpretable sequencing ladder was estimated to be in the range of 20 to 50fmols per reaction. However, the sensitivity of detection was limited only by enzymatically-triggered background luminescence, and not by the level of signal obtained. The nonradioactive methods described have been used successfully in an 8-plex system.

**DISCUSSION**

Although the original multiplex protocol was based on a set of tagged vectors (3), tagged primers have also been used or proposed for various forms of multiplex DNA sequencing (George Church, personal communication; 2, 27). For example, a proposal was recently put forward for multiplex sequencing using sequence-labelled primers and fluorophor-labelled probes (1), similar in principle to the methods used here. However, we use tagged primers and the superposition of the four sequencing reactions to address the problem of reading DNA sequence films; a part of this solution is to utilize M13 dideoxynucleotide sequencing, thereby improving the quality of the data to be analyzed. In addition, the proposal for fluorophor-labelled probes does not take into consideration any of the practical sequencing problems addressed here, and, in the version described, remains a promising but unproven scheme for large scale DNA sequencing.

There are several advantages to tagging primers instead of vectors. Firstly, there is no need to prepare multiple libraries of clones in special vectors. This means that workers can use vector/host combinations that yield good results in their hands, and an increased depth of multiplexing can easily be accommodated by synthesizing more primers. This should make multiplexing more accessible to workers undertaking smaller projects. A theoretical disadvantage of tagged primers is that the procedure can only be multiplexed following primer annealing (1), or following the sequencing reactions (this paper; in practice, pooling immediately after the annealing step might lead to increased backgrounds if one or more primers were present in excess over their template DNAs). This is a relatively late stage. In the original procedure (3), clones were pooled prior to amplification by growth, an early step. However, we do not believe the sacrifice to be of practical importance when using phage vectors. In our experience, recombinant M13 phage have variable growth rates and the effects of competition are likely to severely limit the number of clones that can usefully be pooled for growth. In contrast, by growing clones individually, the depth of multiplexing is only really limited by probe sensitivity. We have not investigated the factors influencing variability in phage growth rates.

It is worth noting that reliable protocols have been developed for growing large numbers of individual M13 clones and preparing high quality ssDNA templates in microtitre trays (28, 21). It is relatively simple to prepare manually two microtitre trays of ssDNA templates (120 clones) in a day. Sufficient clones can be prepared in a week to sequence a 20kb fragment to a redundancy of 10 (Victoria Smith, personal communication). In this laboratory, ssDNA is now prepared with the aid of a commercially available robotic workstation (21). As sequencing reactions are also carried out in microtitre trays, manually or robotically (20, 29), the entire M13-based dideoxynucleotide sequencing procedure is amenable to automation (29). For these reasons we see little practical advantage in pooling clones early. Finally, by not pooling clones early, the ability to easily retrieve individual clones is retained, which may facilitate directed sequencing later in a project should this become necessary.

Multiplex DNA sequencing is currently limited by the lack of a robust computer program which can correct for the large variety of gel and sequencing artefacts that are normally encountered. The foundation of a film reading program is the ability to bring into register precisely vertical arrays of base-specific bands. This requires the ability to track lanes, correct for distortions, and order bands based on their relative spacing. A method of sequencing which has successfully overcome the problem of sequence reading uses real-time detection of fluorescently labelled DNA samples migrating through the gel (2). This system also utilizes the principle of running the four base reactions down the same lane (2). However, bands are detected at a fixed location in space, and their detection is separated in time. Hence the

![Figure 3. Four separate base-specific reactions imaged from a single lane using chemiluminescent detection.](image-url)
REFERENCES

31. Davison, A. DNA Sequence, in press.

ACKNOWLEDGEMENTS

I am particularly grateful to George Church for thought-provoking discussions and gifts of vectors and oligonucleotide probes and to John Sulston for advice. I also thank Victoria Smith for the gift of DNA samples, Bart Barrell for long-term support, Tom O'Keefe of Milligen/Biosearch for lessons in multiplexing and Amersham International for the optical density overlays shown in Figure 2. M.C. is supported by a fellowship from Applied Biosystems.

problem of gel distortion is essentially avoided, although corrections for the different mobilities of the four dyes must be carried out. In contrast, we utilize the advantages of single lane electrophoresis to address the problem of superimposing four relatively large and complex two-dimensional images. Furthermore, by using sequence-tagged oligonucleotides which are detected by hybridization, a much greater depth of multiplexing can realistically be achieved than by real-time detection.

The use of two-dimensional colour traces to depict the processed output of a film reader is consistent with the method of displaying fluorescence traces, and should facilitate the checking and editing of sequence databases in which both kinds of data have been entered. The sequence compilation programs used in this laboratory, which are already capable of handling large shotgun databases (8, 31), have recently undergone extensive improvements (Rodger Staden, personal communication). There is now an interactive database editor which allows the graphical display of fluorescence traces, and it is envisaged that this feature could be extended to allow the handling of data from a film reader when a suitable machine is developed.