Volume 16 Number 7 1988 Nucleic Acids Research

Solid phase DNA sequencing using the biotin–avidin system

Stefan Ståhl, Thomas Hultman, Anders Olsson*, Tomas Moks and Mathias Uhlén+

Department of Biochemistry, Royal Institute of Technology, S-10044 Stockholm, Sweden

Received December 23, 1987; Revised and Accepted February 19, 1988

ABSTRACT

A novel method for solid-phase DNA sequencing is described. A plasmid vector, pRIT27, has been designed to allow directional immobilization of double stranded plasmid to avidin agarose. The strategy involves enzymatic incorporation of 11-bio-dUTP into the plasmid and strand specific elution using alkali. The immobilized single stranded DNA is used as template for sequencing reactions and the resulting labelled oligonucleotides are eluted by alkali. The affinity gel containing the immobilized template is consecutively used for the four different dideoxy-nucleotide reactions. The solid-phase technique can be used for both primer specific or extension specific labelling. The possibility to use the system in automated DNA sequencing is discussed.

INTRODUCTION

The increased interest in large scale sequencing projects, as exemplified by the proposals to sequence the entire human genome (1), necessitates technical improvements to enable megabase sequencing. Sequencing can be divided into six operations, which all have to be automated to facilitate large scale projects. These are (i) template preparation, (ii) sequence reactions, (iii) electrophoresis, (iv) detection of specific fragments, (v) data retrieval and storage and (vi) data analysis. Except for data storage and analysis, most steps are at present carried out manually, causing considerable investments in operator time.

Recently, a number of technical improvements have been reported. Data evaluation, i.e. computer software (2), has been further developed. A number of technical improvements have been described based on the strong interaction between biotin and avidin, for applications such as colorimetric detection of DNA sequencing (3), rapid plasmid screening (4) and oligonucleotide sequencing (5). A filter method to prepare single stranded phage

DNA has also been described (6), which may be used in an automated procedure. Attempts to develop automated sequencing reactions by a centrifugal reagent handling device have also been described (7), as well as image processing programs for the detection of the bands on the autoradiograms (8). However, the most common approach to automate techniques is the use of laboratory robots. Systems for high-speed sequencing (9) and template preparation (10) have been introduced based on this strategy.

A novel approach to automate the electrophoresis step has been described by several groups (11,12,13) where fluorescence instead of isotopes is used for labelling the DNA fragments. These systems enable on-line detection, whereby the three operations electrophoresis, detection and data acquisition may be combined in a single automated step. This approach is therefore likely to be included in megabase sequencing strategies.

To obtain a completely automated sequencing protocol, it is essential to develop methods that can be automated also for the first two operations (template preparation and sequencing reactions). Recently, work stations for automated preparative purification of plasmids have been developed (ref.10 and T. Moks and M. Uhlén, manuscript in preparation). In this paper, we describe a novel strategy for the sequencing operation. This is based on the immobilization of double stranded plasmid DNA to a solid support and subsequent use of the immobilized DNA as template for sequencing reactions. This solid-phase approach facilitates automated handling of liquids in µl quantities and is therefore well adapted for automatization.

MATERIALS AND METHODS

DNA techniques.

Restriction enzymes were obtained from Pharmacia, Sweden. DNA manipulations and purifications were performed according to standard procedures (14) and 11-bio-dUTP was obtained from BRL and was used to terminally labelled double stranded DNA according to the suppliers recommendations. Avidin agarose was obtained from Sigma chemicals.

Construction of plasmid pRIT27.

Plasmid pTZ18R (Pharmacia, Sweden) was digested with BglI

and a synthetic oligonucleotide linker 5'-CCATGACAATGGAGTGCTGG TTACCGATATCGAA-3' (and its complimentary sequence) was inserted. This synthetic fragment contains BstXI, BstEII and EcoRV recognition sequences. The BglI site, used for the insertion, was destroyed simultaneously. The reading frame was changed in the last part of the lacZ'-gene, but the color of the colonies remained light blue when introduced into E.coli strain RRI with IPTG/Xgal selection (14). This construct was digested with EcoRI and HindIII, and a synthetic oligonucleotide linker 5'-AATTCGGCC AGCACGGCCGGCTCAGGTGACCA-3' (and its complementary sequence) was inserted. The EcoRI and HindIII sites were thereby destroyed and a sequence of SfiI, EcoRI, PstI, HindIII and SfiI sites was created. This insertion changed the color of the colonies from blue to white, due to a frame shift in the lacZ' gene. The new EcoRI and HindIII recognition sequences were used to insert a mp8 multilinker, restoring the correct frame in the lacZ' gene, thus giving blue colonies. Thereafter the PvuII site upstream of the lacZ' gene was converted into a BglII site, by insertion of a linker 5'-CAGATCTG-3' (Kabigen, Sweden). The resulting plasmid was denoted pRIT27.

Immobilization of biotinylated double stranded DNA.

Plasmid pRIT27 containing an insert derived from the multilinker region of M13 mp18 (15), was digested with BstEII and EcoRV. The 5' protrusions were filled in with Klenow polymerase (14) using 11-bio-dUTP and appropriate dNTP's. The material was purified using a Sephadex G-50 (Pharmacia, Sweden) column, followed by ethanol precipitation. After redissolving in TE (10 mM TRIS pH 7.5, 1 mM EDTA) the plasmid was digested with EcoRI. This reaction mixture containing the biotinylated double stranded DNA was mixed with avidin agarose gel, prepared by washing with 1 M NaCl and TE. Approximately 1 µg of plasmid (treated as described above) was used per µl avidin agarose gel for the immobilization. The mixture was inverted gently at room temperature for one hour.

Sequencing reactions using immobilized template DNA.

The immobilized biotinylated double stranded DNA was converted into single-stranded form by incubation at 37°C with 0.15 M NaOH for 15 minutes. The avidin agarose gel, with immobilized

template DNA was subsequently washed with 0.15 M NaOH and H_2O. Sequencing reactions were performed using both ^{35}S labelled dATP and ^{32}P end-labelled primer. In both cases 1 µg of the plasmid immobilized on 1 µl avidin agarose gel was mixed with oligonucleotide primer, in a total volume of 2.5 µl. The mixture was incubated at 60°C for 1 hour and allowed to cool at room temperature. 5 µl of the appropriate nucleotide mix was added and the volume was adjusted to 10 µl with a buffer containing 10 mM Tris HCl (pH 7.5) 10 mM $MgCl_2$, 100 µg/ml BSA and 100 mM NaCl. For the ^{35}S protocol, 0.5 µl ^{35}S-dATP (12.5 µCi/µl) was also added. In both cases, the reaction mixtures, in a total volume of 10 µl, were incubated 20 min at 37°C. For the ^{35}S protocol the following nucleotid mixes were used: Amix; 62.5 µM dCTP, dGTP, dTTP, 25 µM ddATP; Cmix; 83 µM dGTP, dTTP; 4 µM dCTP, 50 µM ddCTP; Gmix; 83 µM dCTP, dTTP, 4 µM dGTP, 150 µM ddGTP; Tmix; 83 µM dCTP, dGTP, 4 µM dTTP, 125 µM ddTTP. For the ^{32}P labelled primer the following nucleotide mixes were used: Amix; 83 µM dCTP, dGTP, dTTP, 4 µM dATP, 50 µM ddATP; Cmix; 83 µM dATP, dGTP, dTTP, 4 µM dCTP, 50 µM ddCTP; Gmix; 83 µM dATP, dGTP, dTTP, 4 µM dGTP, 150 µM ddGTP; Tmix; 83 µM dATP, dCTP, dGTP, 4 µM dTTP, 250 µM ddTTP.

After completed reactions the supernatant was removed by centrifugation and the gel washed three times with 50 µl of H_2O. The newly synthesized oligonucleotides were eluted using 10 µl 0.15 M NaOH and the eluant was subsequently neutralized with HAc. The samples were ethanol precipitated and redissolved in 5 µl TE. A fraction of 2 µl was mixed with 2 µl formamid/dye mix and heated for 3 min in boiling water and loaded on a 6% polyacrylamide sequencing gel. The avidin agarose gel, with immobilized template DNA, was regenerated by extensive washing with 0.15 M NaOH and H_2O.

RESULTS

The basic concept

The general principle of the method is outlined in Fig. 1. The target DNA is cloned into the multilinker region of the sequencing vector, pRIT27. The plasmid is linearised with restric-

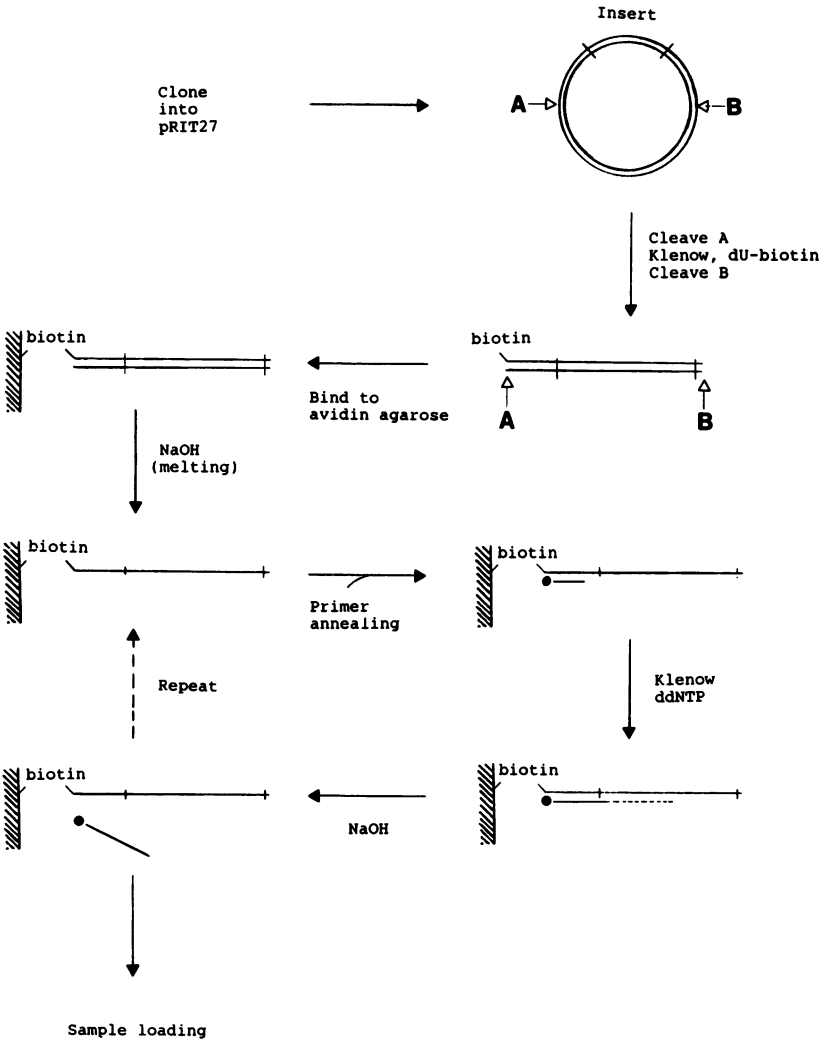

Fig.1. A schematic drawing of the basic concept of the solid-phase sequencing using the biotin-avidin system. Note that depending on the choice of enzymes A and B, both fragments will or will not be immobilized. See text for details.

tion enzyme A and the protrusions are filled in using Klenow polymeras and 11-bio-dUTP, dATP, dCTP and dGTP. After restriction with a second enzyme B, the mixture is passed over an avidin agarose column. This leads to directed immobilization of the bioti-

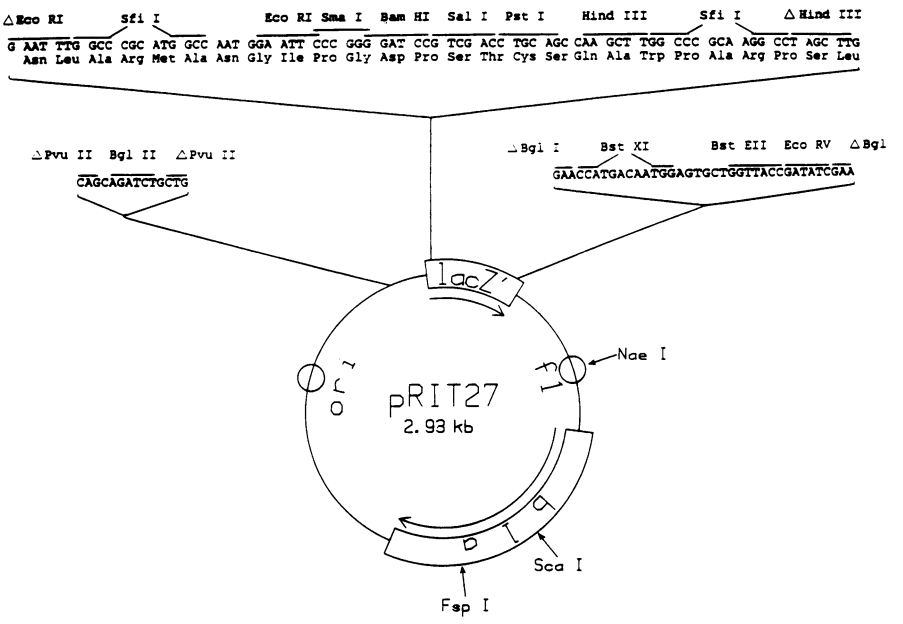

Fig.2. The sequencing vector, pRIT27. The nucleotide sequence and the deduced amino acid sequence in the multi-linker region is shown as well as the sequence of the synthetic linkers inserted in the flanking regions. Abbreviations: bla, ß-lactamase gene; ori, origin of replication; f1, origin of replication of phage f1; lacZ', part of the ß-galactosidase gene.

nylated DNA fragments (Fig. 1). Note that both fragments are immobilized using this procedure. If only the fragment containing the insert is to be immobilized, it is possible to use an enzyme A recognizing a non-palindromic sequence or to use a third enzyme C restricting the vector immediately adjacent to the site recognized by enzyme A.

Single stranded DNA is obtained by melting the strands, either by alkali or heat treatment, and simultaneous elution of the non-biotinylated strand. A general sequencing primer is annealed to the resulting immobilized single stranded template and the sequencing reaction is performed under standard conditions (15). The extended oligonucleotides can be labelled using different strategies, most notably isotopes or fluore- scence which are incorporated either during the extension or as a labelled primer. The newly synthesized labelled oligo- nucleotides are e-

luted by another melting step leaving the template available for the next sequencing reaction. The annealing and extension is repeated to obtain specific fragments for all four nucleotides and the four samples are loaded on a sequencing gel.

The sequencing vector pRIT27

A new multi-purpose plasmid vector pRIT27 was constructed, to enable sequencing using the concept outlined in figure 1. A schematic drawing of the plasmid is shown in figure 2. The plasmid is a high-copy number vector containing an f1 origin of replication, which generates single stranded plasmid when the E.coli host is superinfected with phage M13 (16). This feature is not essential for the solid-phase application, but might be important for other applications, such as in vitro mutagenesis. A multi-linker site in the lacZ' region gives the blue/white selection common for many cloning vehicles (15). Two flanking SfiI--sites were designed to facilitate uni-directional transfer of inserted material to shuttle and expression vectors. For solid-phase sequencing, linkers were inserted approximately 200 base pairs upstream and downstream from the multi-cloning site. Cleavage with BglII or BstEII yields protruding 5'-ends, which can be used to incorporate biotinylated uracil using Klenow polymerase.

The linker containing the BstEII site was inserted into the BglI site in the 3'-end of the lacZ' gene. This affects the ability of the gene product to complement the ß-galactosidase and yields light blue colonies, but there is still a distinct difference between white and light blue colonies (data not shown). Details of the construction is described in Material and Methods.

Immobilization of plasmid DNA to avidin agarose.

To evaluate the effiency of the immobilisation and the following melting step, pRIT27 was digested with BstEII and EcoRV and the 5' protrusions were filled in using 11-bio-dUTP and appropriate dNTP's. Following EcoRI digestion and alkaline phosphatase treatment, the fragment was end-labelled using ^{32}P--dATP and polynucleotide kinase. The latter step allowed us to monitor the efficiency of the procedure by measuring the radioactivity in the various fractions.

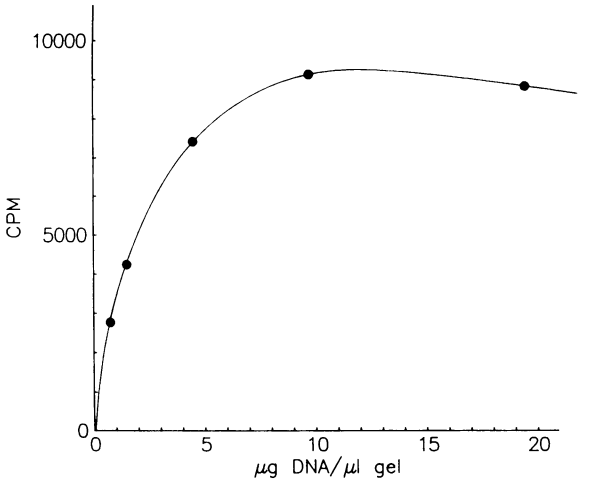

Fig.3. Immobilization of biotinylized, double stranded pRIT27, end-labelled with ^{32}P. The amount of label bound to 1 ul of avidin agarose after 30 minutes of incubation at room temperature is shown. See text for details.

The capacity of the avidin agarose was determined in a saturation experiment. Avidin agarose was mixed with increasing amounts of labelled and biotinylated plasmid DNA and the amount of immobilized labelled material was determined. The result (Fig.3) demonstrates that several μg of DNA can be bound to each μl of avidin agarose, suggesting that the capacity of the matrix is sufficiently high to allow sequencing reactions on a reasonable scale.

Elution of complementary strand using alkali.

To test if the immobilized double stranded plasmid DNA could be converted into a single-stranded form by strand specific elution with alkali, different amounts of biotinylated and ^{32}P-labelled plasmid DNA was bound to 1 ul of avidin agarose. After extensive washing with TE, the gel was incubated at 37°C with 0.1 M NaOH for 10 minutes. The amount of labelled DNA bound to the gel, before and after the alkali treatment, was determined. The results presented in Table 1 show that almost all of the label could be immobilized. Denaturation and elution of the complementary strand released more than 40 percent of the total immobilized label (Table 1), close to the expected value (80 per-

Table 1. Immobilization of biotinylated pRIT27 to avidin agarose followed by strand specific elution.

Amount pRIT27 (µg)	Effiency (%)	
	Immobilization	Elution
0.3	97	ND
0.6	98	42
1.5	96	43

Various amounts of labelled pRIT27 were bound to 20 µl of avidin agarose and the amount of label in the supernatant (not bound) was determined. The gel was thereafter treated with 0.15 M NaOH and the supernatant was collected and measured to yield eluted label. The efficiencies of immobilization and elution were estimated from these results and are presented as the percent of total label. ND is not determined.

cent). These results demonstrate that the biotin-avidin affinity system can be used with high efficiency for immobilization of plasmid DNA followed by strand-specific elution.

Solid-phase sequencing using labelled deoxynucleotide.

The sequencing reactions were performed using a protocol involving labelling of the specific fragments with ^{35}S during extention. A nucleotide mix was used containing S^{35}-labelled dATP and one of the dideoxynucleotides, in addition to the standard nucleotides. In this and the following experiments, a plasmid was used, consisting of pRIT27 containing an insert derived from the multi-linker region of M13 mp18.

Approximately 1 µg of the plasmid was used for the immobilization to 1 µl avidin agarose and the subsequent strand specific elution was performed as described above. A RIT primer (17), complementatary to a region immediately downstream from the multi-linker region, was used for initiation. Equimolar amounts of immobilized single stranded DNA and RIT primer were mixed, incubated for 1 hour at 60°C and allowed to cool to room temperature. The supernatant was removed and the primer extention was started with the appropriate nucleotide mix in a total volume of 10 µl, followed by a chase reaction (15) to extend fragments not terminated with a dideoxynucleotides. After completion of the reactions, the supernatant was removed and the gel extensively washed. The newly synthesized oligonucleotides

were then eluted using 0.15 M NaOH and the eluant was neutralized with HAc. The affinity gel containing the single stranded template was thereafter used for another round of sequencing

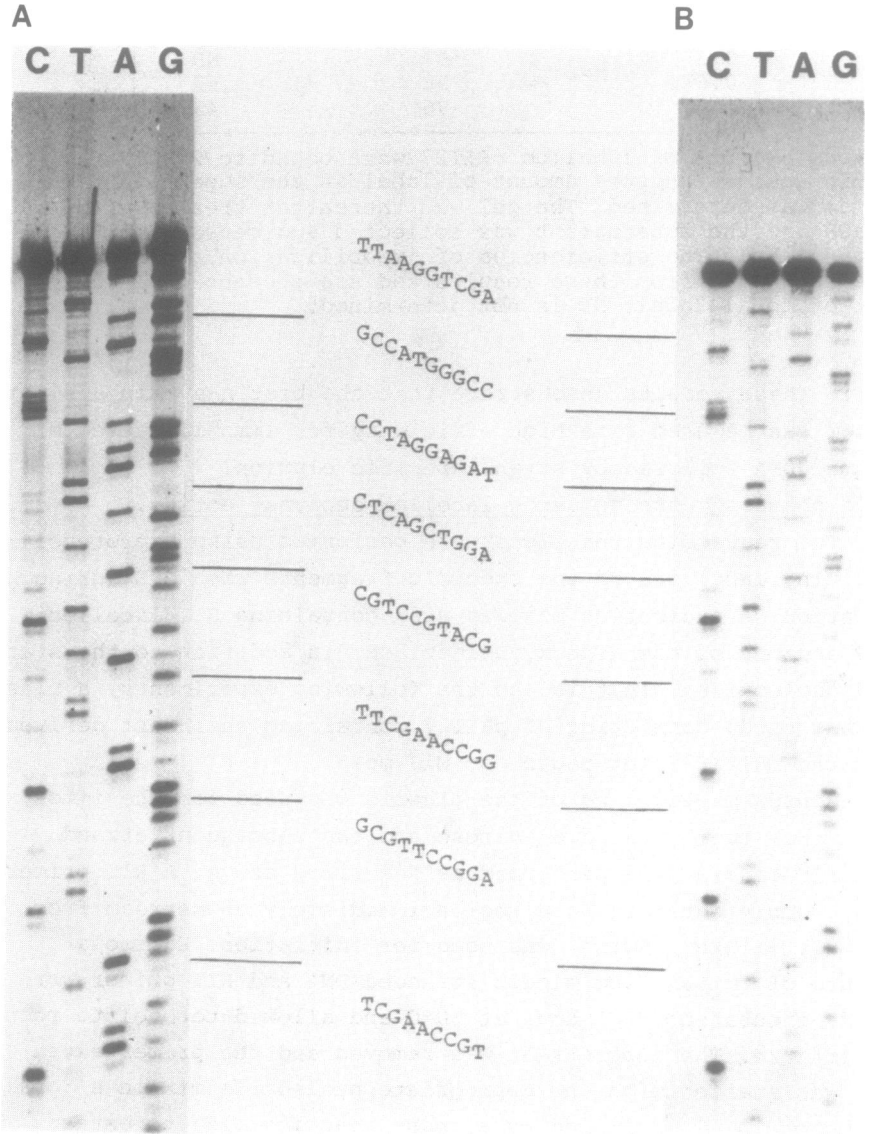

Fig.4. Autoradiographs of sequencing gels with samples obtained by solid-phase sequencing. A; labelling using ^{35}S-dATP during the extension. B; labelling using ^{32}P end-labelled sequencing primer. Also shown is the expected sequence.

reactions, involving primer annealing followed by extention using a new dideoxynucleotide mix.

The protocol was followed for all four dideoxy nucleotides and the eluted samples were ethanol precipitated and redissolved in formamide/dye mix prior to loading on a sequencing gel. An autoradiogram of DNA fragments separated by electrophoresis is presented in Fig.4A. Clearly readable sequences are obtained, which correlates well with the expected sequence of the plasmid used. The strong band at the top of the sequence represents run off transcripts at the EcoRI site at the the 5' end of the immobilized template.

Solid-phase sequencing with end-labelled primer

An alternative strategy was also tested where ^{32}P end--labelled RIT primer was used to label the extended DNA fragments. A similar protocol was used, with the composition of the nucleotide mixes appropriately adjusted (see Materials and Methods for details). Different molar ratios of immobilized template and primer were tested. With equimolar amounts of template and primer (Fig. 4B) a clear and easily readable autoradiogram was obtained. Similar results were obtained for other primer/template ratios (data not shown), suggesting that the ratio can be varied within certain limits without critically influencing the pattern.

DISCUSSION

A procedure for solid phase sequencing using the enzymatic method to generate base specific fragments is described. The immobilization procedure involves enzymatic incorporation of 11-bio-dUTP and binding of the biotinylated fragment to avidin agarose followed by strand specific elution. There are several advantages with this sequencing strategy. First, compared to standard protocols involving single stranded templates, this procedure uses a double stranded fragment as starting material. This facilitates subcloning and makes the system versatile since sequencing can be performed directly in multi-purpose vectors for expression or alternative host range. Second, using the restriction sites BglII and BstEII and two alternative primers, the insert can be sequenced from both directions, which doubles the

amount of information collected from each clone. Although the sequence vector pRIT27 only contains these two recognition sites, it is possible to make the system even more versatile by incorporating additional sites recognized by "rare cutters". Third, the template is reusable, which makes it possible, if necessary, to amplify the specific fragments by repeated procedures. It is also possible to reuse the template for sequence reactions designed to obtained longer extentions, i.e. with different compositions of the nucleotide mixes. In addition, the same template can easily be used for "primer walking", in which the sequence information obtained for the first extension is used to synthesize a new primer for the second extension. This can be repeated for several rounds and large inserts can be sequenced, without subcloning or restriction mapping. The fact that the procedure can be performed on both strands simultaneously, makes it possible to pursue such sequencing with considerable speed. Finally and maybe most important, the directed immobilization on a solid support provides a system which potentially can be automated without requiring mechanical operations, such as centrifugations and pipetting.

The results presented in figure 4 demonstrate that the solid-phase technique can be used for either extention or primer specific labelling. The former way of incorporating label into the specific fragments is the normal way of sequencing at present, while the latter method is used in the newly developed strategies involving fluorescent dyes (11,12). Although both methods were found to give satisfactory results, we have found less ambiguities with primer specific labelling (Fig. 4B). This is most likely due to non-specific priming when the extention labelling is used, although optimization of the conditions, such as the use of T7 polymerase, might limit the appearance of artifactual bands.

The matrix used was avidin agarose, but other types of supports may be considered depending on requirements such as regenerability, volume limitations and hardware configurations. Magnetic beads might be advantageous for certain applications, while solid supports such as membranes or coated capillaries, might be desired for other applications. Thus, although the same

general concept (Fig.1) may be applied, the choice of solid support can be flexible.

The protocol described here has been adapted to facilitate automation. However, suitable manual procedures can also be envisioned where the annealed primer-template matrix is divided into four tubes and the extention reactions are run in parallel. Such procedures using magnetic beads as solid support has recently been found to give clear and reproducible results (T.Hultman and M.Uhlén, unpublished).

Although we have only shown that the solid-phase strategy outlined in figure 1 can be used for DNA sequencing, other applications requiring immobilized single or double stranded DNA may also be considered. This includes cDNA synthesis, selection or enrichment of specific RNA sequences and affinity purification of DNA binding proteins.

It is important to evaluate if the solid-phase strategy, outlined in figure 1, can be incorporated into a general sequencing strategy. Together with an automated electrophoresis station (11,12,13) and an automated station for preparative plasmid purifications, the solid-phase strategy may provide a completely automated and consecutive scheme for all operations in sequencing. This will enable sequencing in considerably larger scale than is possible at present.

ACKNOWLEDGEMENT

This work was supported by grants from the Swedish Board for Technical Development. We are grateful to Drs Lennart Philipson, Lars Abrahmsén, Andras Gaal and Sven-Erik Sköld for critical comments and advice and to Gerd Benson for typing this manuscript.

*Present address: Pharmacia-LKB Biotechnology, Bromma, Sweden

+To whom correspondence should be addressed

REFERENCES

1. Smith,L. and Hood,L. (1987) Biotechnology 5, 933-939
2. Smith,D.H., Brutlag,D., Friedland,P. and Kedes,L.H. Nucl.Acids Res. (1986) 14, 17-20
3. Beck,S. (1987) Anal.Biochem. 164, 514-520.

4. Rigas,B., Welcher, A.A., Ward,D.C. and Weissman, S.M. (1986) Proc.Natl.Acad.Sci.USA 83, 9591-9595.
5. Rosenthal,A., Schwertner,S., Hahn,V. and Hunger,H.-D. (1987) Nucl.Acids Res. 13, 1173-1184.
6. Kristensen,T., Voss,H. and Ansorge,W. (1987) Nucl.Acids Res. 15, 5507-5515
7. Martin,W.J., Warmington,J.R., Galinski,B.R., Gallagher,M., Davies,R.W., Beck,M.S. and Oliver,S.G. (1985) Biotechnology 3, 911-915
8. Elder,J.K., Green,D.K. and Southern,E.M. (1986) Nucl.Acids Res. 14, 417-425
9. Wada,A., Yamamoto,M. and Soeda,E. (1983) Rev.Sci.Instru. 54, 1569-1572
10. DeBonville,D.A. and G.E.Riedel. (1987) In Advances in Laboratory Automation Robotics. 1986. J.R.Strimaitis and G.L.Hawk, eds. Zymark Corp. Hopkinton, MA. pp. 353-360
11. Smith,L.M., Sanders,J.Z., Kaiser,R.J., Hughes,P., Dodd,C., Connel,C.R., Heiner,C., Kent,S.B.H. and Hood,L. (1986) Nature 321, 674-679
12. Ansorge,W., Sproat,B., Stegemann,J., Schwager,C. and Zenke,M. (1987) Nucl.Acids Res. 15, 4593-4602
13. Prober,J.M., Trainor,G.L., Dam,R.J., Hobbs,F.W., Robertson,C.W., Zagursky,R.J., Cocuzza,A.J., Jensen,M.A. and Baumeister,K. (1987) Science 238, 336-341
14. Maniatis,T., Fritsch,E.F. and Sambrook,J. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, NY, 1982
15. Messing,J. (1983) Meth.Enzymol. 101, 20-39
16. Dente,L., Cesarini,G. and Cortese,R. (1983) Nucl.Acids Res. 11, 1645-1655
17. Olsson,A. and Uhlén,M. (1986) Gene 45, 175-181
15. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc.Natl. Acad.Sci.USA 74, 5463-5467