Volume 8 Number 23 1980 Nucleic Acids Research

A subcloning strategy for DNA sequence analysis

A.M.Frischauf, H.Garoff and H.Lehrach

European Molecular Biology Laboratory, Postfach 102209, 6900 Heidelberg, GFR

Received 20 October 1980

SUMMARY

We describe here a new strategy of fragment preparation for sequencing procedures using endlabelled DNA fragments as substrates (2,3) which is directly applicable to DNA fragments cloned into the Pst I site of pBR322, or in modified form, to inserts into the BamH I or Sal I site of the same plasmid. Ordered sets of subclones of predetermined overlap are generated. These can be sequenced directly without further strand- or fragment separation steps.

INTRODUCTION

Fast and reliable DNA sequencing techniques (1,2,3) have become available during the last years allowing the determination of nucleic acid sequences of up to 400 bases in length (4) from one end of a DNA molecule. Sequences longer than this have to be determined in shorter stretches and then assembled to give the entire nucleic acid sequence. Selection of fragments for sequencing has either used a deterministic strategy, usually based on a preexisting restriction map or involved a random selection procedure, often relying on computer programs to identify and assemble overlapping sequences (5,6).

Both of these strategies have some inherent difficulties. The establishment of detailed restriction maps of long DNA molecules can involve considerable effort and in some cases, possibly due to unusual features of the DNA sequence, no convenient restriction sites for sequencing can be found.

Random sequencing strategies, in contrast, will in general require a large number of sequencing experiments to ensure the complete representation of the entire DNA in the assembled sequences.

We describe here a strategy which allows essentially sequence and restriction site independent non-stochastic sequencing. In addition, similar to a sequencing strategy using subcloning into single stranded phage vectors (7), the technique described here does not involve any gel purification and extraction steps in preparing the samples for sequencing.

MATERIALS

Eco RI linkers were obtained from Collaborative Research, DNAase I, Cla I and Hind III from Boehringer Mannheim, DNA ligase was a gift from Vince Pirrotta, DNA polymerase from W. McClure. T4 polynucleotide kinase was isolated as described by Kleppe et al. (8). Low Melting Point agarose was purchased from Bethesda Research Laboratories.

METHODS

5 µg plasmid were digested with 400 pg DNAase I in 60 µl 0.02 M Tris pH 7.4 1.5 mM $MnCl_2$ for 15 minutes at 25°C . The sample was then extracted twice with phenol:chloroform:isoamylalcohol 50:48:2, 0.1 volume 3 M sodium acetate pH 6 and 2.5 volumes ethanol were added and the sample was precipitated in dry ice. This preparation was treated with E. coli DNA polymerase I to polish the ends and ligated to radioactively labelled Eco RI linker exactly as described by Heffron et al. (9). After cleavage with Eco RI nuclease 2 equivalents of disodium ethylenediaminetetraacetic acid were added, the sample was phenol extracted and applied to a 0.8% low melting point agarose gel in 40 mM Tris 5 mM sodium acetate disodium, 1 mM ethylenediaminetetraacetic acid pH 7.6. Approximately 30 slices were taken covering the sizes from the large Eco RI-Pst I fragment of pBR322 and linearised pBRSFV2. The slices of approximately 25 µl volume were diluted with 200 µl water, heated to 70°C for 10 minutes, cooled to 37°C, 25 µl 0.2 M Tris/Cl pH 7.6, 0.1 M $MgCl_2$, 10 mM dithiothreitol and 50 units/ml T4 DNA ligase were added and the mixture incubated overnight at 15°C. After vortexing the tubes hard for a few seconds to break up possibly reformed gel matrix, aliquots of the samples were then used to transform E. coli X1776 according to a modification of a procedure developed by Doug Hanahan (10). Clones were picked at random among the transformants from successive slices and plasmid was isolated. The purified DNAs were then cleaved with Eco RI, and either labelled at the 5' end by sequential treatment with bacterial alkaline phosphatase and T4 induced polynucleotide kinase and γ^{32}P-ATP or at the 3' end by incubation with α^{32}P-dATP and the large fragment of DNA polymerase I. To separate the labelled ends the DNA was recleaved with Cla I or Hind III. The short labelled fragment was either left in the sequencing reaction or removed by agarose chromatography. Materials and conditions for these enzymatic reactions and the sequence analysis have been described (11,12).

RESULTS

Figure 1 illustrates the principal steps in the construction of the ordered set of subclones to be used for the sequencing reactions. We first introduce a

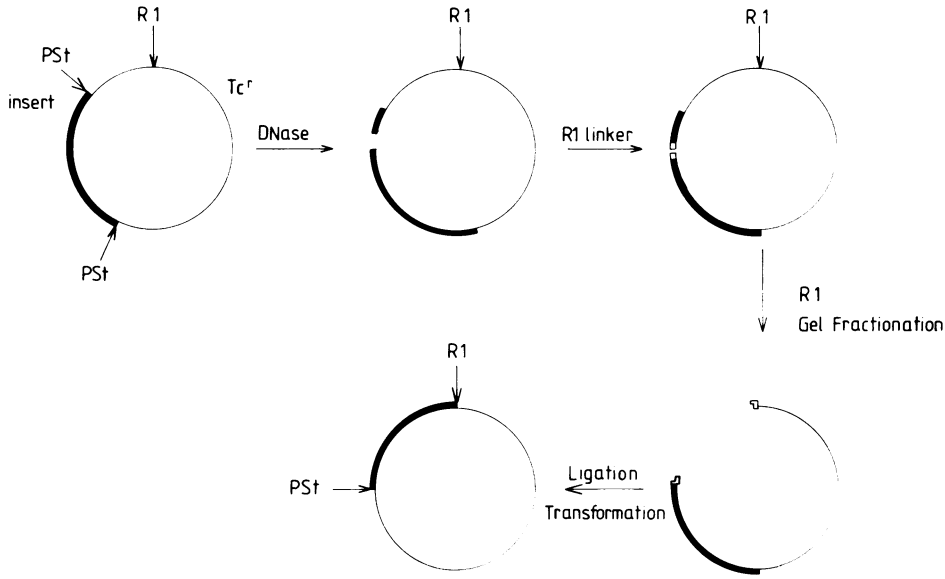

Fig. 1: Scheme of subclone construction.

moderate amount of double strand cuts (approximately 50% of the molecules are cleaved to ensure a low level of multiply cleaved molecules) by a limited treatment with pancreatic DNAase I in a manganese buffer system (9) (figure 2).

After a short treatment with DNA polymerase I to increase the proportion of flush ends we then attach chemically synthesised linker molecules carrying Eco RI restriction sites to the cleavage points. We then use Eco RI to cleave both attached linker molecules and the Eco RI site carried by the vector. Originally uncut molecules will simply be linearised by this treatment while the other molecules will be cut into two fragments, the length of which will depend on the position of the DNAase cleavage point. This mixture is then fractionated on a low melting point agarose gel to separate the population of DNA molecules according to size. A region of the gel containing the size classes created by cleavages inside the cloned DNA fragment is then divided into slices containing DNA fractions which differ by a predetermined amount (ideally the length of the DNA sequence readable from one end minus the required overlap) in length. Each fraction will contain predominantly two types of molecules, one extending from the Eco RI site of pBR322, clockwise in our drawing, to a position in the insert, the other one counterclockwise an identical distance to a point located symmetrically in the tetracycline resis-

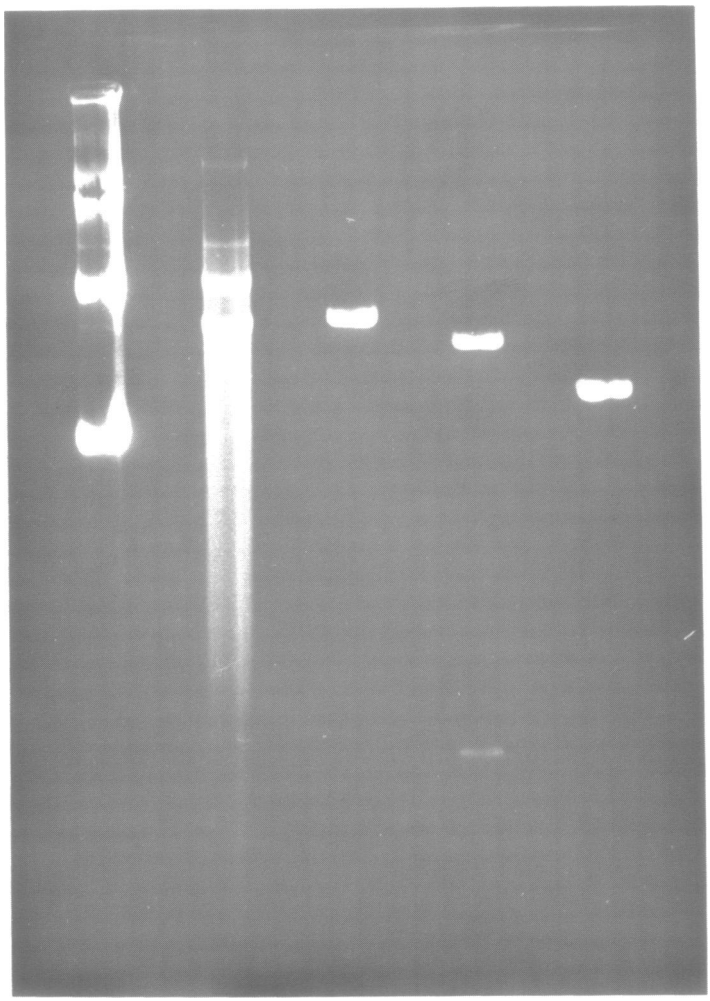

Fig. 2: DNAase digestion of pBR SFV2 on an agarose gel. From left to right: pBR SFV2 before DNAase treatment; pBR SFV2 cut by DNAase; pBR SFV2 cut by Bam HI; pBR SFV2 cut by Eco RI; pBR322 cut by Hind III.

tance gene of the plasmid. This second class of molecules will therefore not be able to transform E. coli cells to tetracycline resistance and will not be found in the following cloning step. After melting the gel slice, the DNA is circularised by ligation, and directly used for transformation, selecting for tetracycline resistance. Individual transformants generated from DNA fractions

of increasing length are picked, plasmid DNA is isolated and sequenced after cleavage with Eco RI, labelling and recleavage with Cla I or Hind III.

This approach was tested in a subcloning experiment on the Semliki Forest Virus cDNA clone pBR SFV2 carrying an insert of approximately 2250 base pairs length in the Pst I site of pBR322 (12). The insert contains an Eco RI site approximately 90 bases from the end close to the position of Eco RI site of the vector. We therefore did not expect to find subclones reaching beyond the Eco RI site of the insert. After subcloning as described above we randomly selected a single clone from each fraction and mapped the extent of the generated deletion by digestion of the plasmid with Ava I. Fig. 3 shows the position of the Ava I and Pst I sites of the parent plasmid pBR SFV2. As shown in Fig. 4, creating deletions of increasing length extending from the Eco RI site in the vector to positions in the insert will first lead to a shortening of the Ava I fragment 2 of the parent plasmid (subclones 1 and 2). Between subclone 2 and 3 one Ava I site is deleted, and the fragment resulting from the deletion can be observed to become progressively smaller. The deletion carried by subclone 10 also removes the other Ava I site in the Semliki sequence and the remaining restriction fragment, linearised by cleavage at the Ava I site in the vector molecule, again decreases further in size.

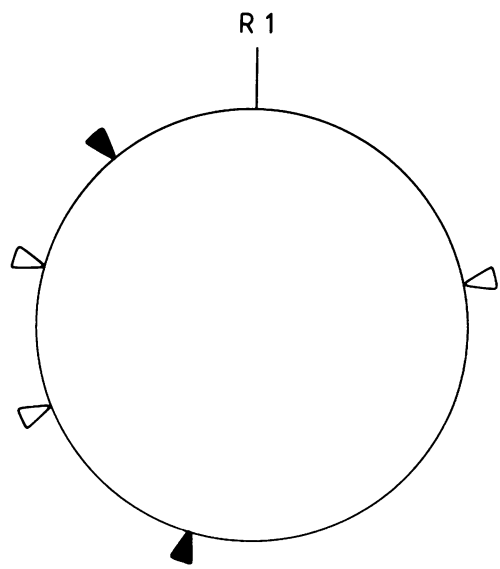

Fig. 3: Map of restriction sites of pBR SFV2 ▽Ava I, ▼Pst I.

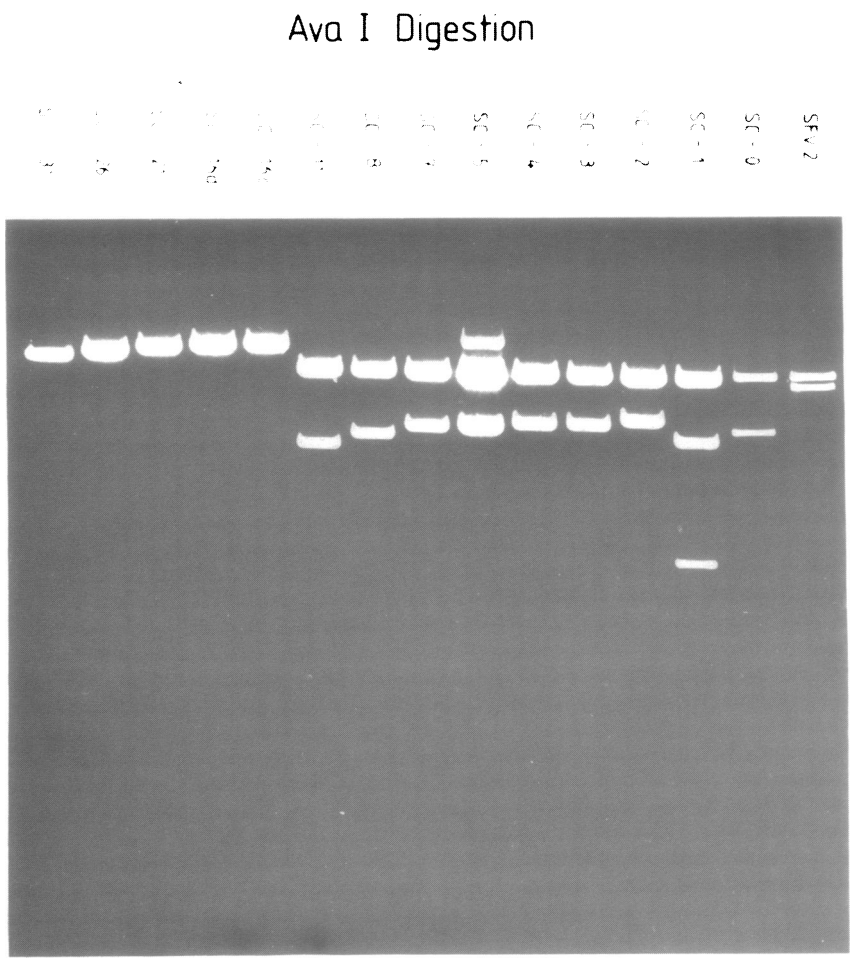

Fig. 4: 1% agarose gel of Ava I digestion of the subclones. Sc 31 is the smallest, SC 0 is the largest subclone. Numbers correspond to gel slices.

DNA from a series of such subclones was then cleaved with Eco RI, labelled using either polynucleotide kinase or DNA polymerase I, recleaved with Cla I or Hind III and in most cases directly used for the Maxam-Gilbert sequencing reactions. Since in this case the sequence of the short, 20 to 30 base pair fragment generated by the recleavage reaction will be superimposed on the beginning of the sequence of the main fragment the first few bases are difficult or impossible to read. Alternatively the short labelled fragments can be removed by an agarose sizing column or by gel electrophoresis.

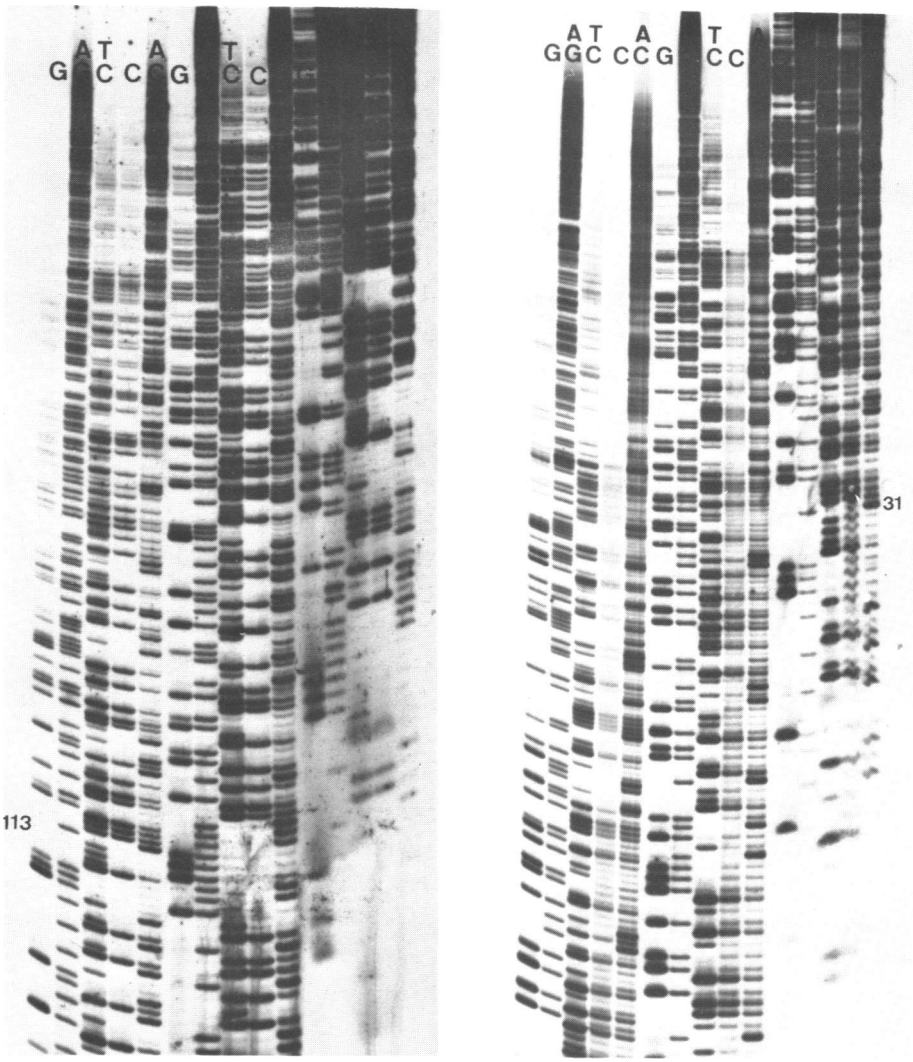

Fig. 5: DNA sequencing gel of subclone 11 (Fig. 5a) and subclone 13 (Fig. 5b). The DNA was endlabelled as described and subjected to base specific cleavage reactions (G, A+G, C/T, C, C/A) and analyzed on a 0.2 mm thick 8% acrylamide gel (H. Garoff, unpublished). The sequence data show that subclone 11 is 82 bases longer than subclone 13. The 31st nucleotide from the labelled end of subclone 13 is indicated and corresponds to the 113th nucleotide (indicated) of subclone 11.

The result of sequencing experiments on subclones 11 and 13 are shown in Fig. 5. Sequences derived by this procedure have been shown to be both consistent between different subclones, and where available, with sequences derived by the direct sequencing of restriction fragments.

CONCLUSIONS

We are describing a technique to construct a series of subclones of increasing length from plasmids carrying inserts in the Pst I site of pBR322, which can be used easily for the sequencing procedures described by Maxam and Gilbert or Maat and Smith. The steps involved in the construction are simple, reliable and involve a minimum of manipulations. In our hands we have had more consistent, and usually better results in sequencing experiments involving subclones than in experiments involving the usual steps of gel separation and gel elution to purify specific DNA fragments for the sequencing reactions. During the sequencing of cloned Semliki Forest Virus 26S mRNA we have found this approach especially useful to cover regions which, due to the absence of restriction sites, have proven to be difficult to cover by any other technique. In addition, by preserving the information on the order of the fragments, potential ambiguities in the assembly of sequences by overlap information are avoided. Though described here for the sequencing of clones containing inserts in the Pst I site of pBR322, this procedure can easily be adapted to clones carrying inserts in the Bam HI or Sal I site of pBR322, using Hind III linkers during the construction.

If Cla I linkers are available, inserts in the Hind III or Eco RI sites can be subcloned analogously. Cla I linkers could also similarly be used as an alternative to Hind III linkers for inserts carrying internal Hind III sites for fragments cloned into the Bam HI and Sal I sites or any combinations of these sites in pBR322. The steps described can similarly be used to construct series of subclones of increasing length for the mapping of genetic function in procaryotic or eucaryotic transoformation experiments. Clones identified as interesting can be directly sequenced using the procedure described above.

ACKNOWLEDGEMENTS

We want to thank Evelyn Kiko and Annemarie Poustka for excellent technical assistance, Will McClure and Vince Pirrotta for enzymes, and Kai Simons for support and discussions.

REFERENCES

1. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

2. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
3. Maat, J. and Smith, A. (1978) Nucl. Acids Res. 5, 4537-4545
4. Sanger, F. and Coulson, A.R. (1978) FEBS Letters 87, 107-110
5. Staden, R. (1979) Nucl. Acids Res. 6, 2601-2610
6. Gingeras, T.R., Milazzo, T.J.P., Sciaky, D. and Roberts, R.J. (1979) Nucl. Acids Res. 7, 529-545
7. Schreier, P.H. and Cortese, R. (1979) J. Mol. Biol. 129, 169-172
8. Panet, A., Van de Sande, J.H., Khorana, P.C., Raae, A.J., Lillehang, J.R. and Kleppe, K. (1973) Biochemistry 12, 5045-5050
9. Heffron, F., So, M. and McCarthy, B.J. (1978) Proc. Natl. Acad. Sci. USA 75, 6012-6016
10. Hanahan,D., personal communication.
11. Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H. and Delius, H. Proc. Natl. Acad. Sci. USA, in press.
12. Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H. and Delius, H. Nature, in press.