Sequencing long DNA fragments cloned in bacteriophage M13 by using internal primers

The sequence analysis of a yeast DNA fragment containing a replication origin

John HINDLEY and Geraldine A. PHEAR

Department of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, U.K.

(Received 26 August 1981/Accepted 17 September 1981)

In the 'shotgun' procedure for sequencing DNA, DNA fragments are cloned into a phage M13 vector and sequenced by using a flanking primer. In a variation of this procedure a longer DNA sequence is cloned into M13, the two single-stranded recombinants identified and sequenced by using a set of internal primers prepared by exonuclease III digestion of restriction fragments.

Among the primed-synthesis DNA sequencing methods the procedure developed by Sanger *et al.* (1980) has become the most widely adopted. An essential feature of this approach is the reduction of the DNA into sets of restriction fragments which are then cloned into the replicative form of phage M13mp2, or one of its derivatives, and the single-stranded recombinant progeny DNAs used as templates in a primed synthesis reaction using a flanking primer which anneals specifically adjacent to the insertion site. This procedure, despite its flexibility and power, can have some drawbacks and we have explored an alternative approach in which, instead of randomly cloning the small restriction fragments, much longer single DNA fragments (2000 nucleotides) were cloned into the M13 vector to produce two sets of single-stranded recombinants corresponding to the insertion of either one or the other of the two complementary strands of the DNA. The sequencing primers were generated by exonuclease III digestion of restriction fragments (Smith, 1979), purified from digests of the original double-stranded DNA fragments. The approach described is similar in concept to the methods used by Sanger *et al.* (1977) and Barrell *et al.* (1979) but extended to take advantage of the M13 cloning technology (Schreier & Cortese, 1979). The application of this procedure to the sequencing of a 1578 base-pair fragment containing a yeast replication origin is described.

Experimental

Preparation and properties of plasmid pDAMY14

Plasmid pDAMY14 was a gift from Dr. D. Beach and was constructed by the insertion of a DNA

fragment, derived from a partial *Sau3A* digest of total yeast DNA, into the single *Bam*HI site of pDAM1 (Beach *et al.*, 1980). Briefly, pDAM1 consists of the bacterial plasmid pBR325 into which a 1500 base-pair *Hae*II fragment containing the yeast *LEU 2* gene (Beggs, 1978) has been inserted into the *Hae*II site of the tetracycline-resistance gene of pBR325 adjacent to the *BAM*HI locus. While pDAM1 can efficiently transform a leucine-requiring strain of *Escherichia coli* (JA221) to *leu*⁺ it lacks a functional eukaryotic replication origin; its ability to transform yeast protoplasts (*Saccharomyces cerevisiae* M16, *leu* 2–3) to the *leu*⁺ phenotype is severely impaired in the sense that, while the transformation frequency is about 2×10^4 colonies per μ g of DNA, the transformants grow extremely slowly yielding barely detectable microcolonies after 5 days incubation on leucine-free minimal plates. In contrast, transformation with DNA from pDAMY14 occurs with a similar frequency but yields cells which grow rapidly in the absence of leucine giving clear colonies after 2 days incubation. On the basis of these observations (D. Beach, unpublished work) it was concluded that the *Sau3A*-derived fragment contained a functional origin which permitted the replication of the plasmid in the transformed yeast cells.

The plasmid was propagated in *E. coli* JA221 *recA1, leuB6, trpΔES, hsdR⁻ hsdM⁺ lacYC600* grown in L-broth supplemented with chloramphenicol (25 μ g/ml) and isolated from stationary phase cells by using the cleared lysate procedure of Clewell & Helinski (1969) followed by banding in a CsCl/ethidium bromide density gradient (Konings, 1980).

Isolation of DNA fragments for cloning and sequencing

Restriction endonucleases were obtained from Bethesda Research Laboratories Inc. and used according to the manufacturer's specification. Restriction fragments were resolved either by electrophoresis on 1 or 2% (w/v) agarose gels or on 8% (w/v) polyacrylamide-slab gels. Bands were located by staining with ethidium bromide and examination under u.v. light. DNA was extracted from agarose gels by using the procedure of Yang *et al.* (1980) and from acrylamide gels by the method of Maxam & Gilbert (1980).

Cloning restriction fragments into phage M13mp2

Sau3A restriction fragments were cloned into the *Bam*HI site of the double-stranded bacteriophage M13mp2/*Bam*HI (Messing *et al.*, 1981). Following transfection into competent *E. coli* strain 71:18 (Gronenborn & Messing, 1978) the mixture was plated out in the presence of β -galactosidase inducer, isopropyl- β -D-thiogalactopyranoside, and the chromogenic substrate 5-bromo-4-chloro-3-indoyl- β -galactoside (Hollenberg *et al.*, 1976). Colourless plaques were picked, transferred individually into 5 ml of medium and grown with shaking for 8–12 h at 37°C. Recombinant phage was isolated from the cell-free supernatant by precipitation with poly(ethylene glycol) 6000 and NaCl and single-stranded phage DNA was prepared by phenol extraction (Sanger *et al.*, 1980).

DNA sequencing

(a) *Primed synthesis from a flanking primer.* DNA sequences were determined by using the dideoxynucleotide chain termination procedure. The sequences at the end of the cloned fragments were determined by hybridizing the 96 base-pair 'universal' primer, derived from M13mp2962 (Heidecker *et al.*, 1980) to the recombinant single-stranded phage DNAs and extending the primer by the usual procedure (Sanger *et al.*, 1980). In most experiments the primer was pretreated with exonuclease III (Smith, 1979) in order to produce essentially single-stranded primer which permitted accurate reading of the sequence through the junction between the M13 sequence and that of the cloned *Sau3A* fragment.

(b) *Primed synthesis from internally derived primers.* A set of six to ten colourless plaques was picked at random from a cloning experiment in which the 1600 base-pair *Sau3A* fragment, derived from pDAMY14, had been ligated into M13mp2/*Bam*HI. Single-stranded DNAs isolated from these recombinants were annealed to exonuclease III-treated flanking primer (as above) and the sequence corresponding to the first 200–250 nucleotides from the priming site was determined. Two different sets

of sequences were obtained corresponding to the two different orientations in which the *Sau3A* fragment had been ligated into the vector. Four of these recombinants were chosen as templates; L1 and L4 gave the sequence from the end of one of the complementary strands and L2 and L3 gave the sequence from the other. The internal primers were prepared by digestion of the 1600 base-pair *Sau3A* fragment with *Alu*I. Of the nine fragments produced, seven were well resolved by electrophoresis on an 8% (w/v) polyacrylamide gel and were eluted and digested with exonuclease III. Each fragment was finally annealed to samples of the two different single-stranded templates and sequences were determined in the usual manner (Sanger *et al.*, 1980).

In an additional set of experiments doubtful regions of the sequence were checked by priming with three *Hae*III fragments similarly isolated from a digest of the *Sau3A* fragment and predigested with exonuclease III.

Results and discussion

Fig. 1 shows a line diagram of a 1578-nucleotide sequence derived from a *Sau3A* digest of the recombinant plasmid pDAMY14 containing a yeast chromosomal replication origin. The entire sequence was determined by using a set of seven *Alu*I and three *Hae*III fragments as primers on the individual full length single strands of the *Sau3A* fragment cloned into the M13mp2/*Bam*HI vector. The sequences at the 3'-ends of the M13 clones were also determined by using the M13 flanking primer and the overall sequence was checked by subsequently cloning the *Alu*I fragments into M13mp2 by using *Eco*RI linkers and sequencing with the M13 flanking primer in the conventional way.

The method is, in theory, capable of independently sequencing all of each strand of duplex DNA. To obtain complete coverage of both strands however, does mean that most of the restriction fragments from a suitable digest need to be used as primers. Since a single gel electrophoresis will not always clearly resolve all such fragments a second restriction enzyme digest is usually required to obtain a further set of priming fragments. We have found that this is usually sufficient to sequence most, if not all, of each strand.

There are three features of the procedure which make this possible.

(i) The sequences derived can be immediately ascribed to one or other of the complementary strands since this depends only on which of the two templates is used for the primed synthesis. In the random-cloning method there is no immediate way of telling which strand of the cloned fragment has been sequenced.

(ii) The ability to sequence in both directions from

Fig. 1. Diagram of 1578 base-pair *Sau3A* fragment showing sequences determined by using a set of seven exonuclease III-digested *AluI* fragments and three *HaeIII* fragments as primers

↓, *AluI* cleavage sites; ↑, *HaeIII* cleavage sites. Numbers immediately above the line show the *AluI* fragments and below, the *HaeIII* fragments. The arrowed sequences are designated such that, for example, L3.4 is the sequence determined by priming template L3 with exonuclease III-digested *AluI* fragment 4. The terminal sequences B2 and P12 were determined by using a flanking primer. The advantages of this procedure (see the text) are: (a) the automatic assignment of each sequence to one or other strand depending only on which template is used. The identical templates L1 and L4 give sequences reading left to right and correspond to one of the complementary strands of the *Sau3A* fragment. Templates L2 and L3 similarly give sequence reading right to left corresponding to the other strand. (b) Each pair of primings, e.g. L4.4 and L3.4, L4.2 and L3.2 etc., give both the sequence of the primer itself, due to the non-processive action of exonuclease III, and extended sequences into both adjacent fragments. The order of fragments is thus automatically determined. (c) The sequence of region 228–575, which contains $5\frac{1}{2}$ direct repeats, is determined directly by read through and its boundaries defined. Confirmation of the sequence around this region was obtained by separately cloning and sequencing *AluI* fragments 1C and 6.

a particular priming site (by annealing the same primer to the two complementary templates) gives the sequences adjacent to each restriction site, thus automatically ordering the different restriction fragments within the cloned sequence.

(iii) The use of exonuclease III-treated primers usually permits the entire sequence of the primer itself to be deduced from the sequences obtained using the two complementary templates. This is because exonuclease III is a non-processive enzyme and the single strands generated by its action are sufficiently variable in length to give overlapping sequences corresponding to the centre of the primer when used to prime the two different templates. As shown in Fig. 1, the sequence generated by priming with seven different exonuclease III-digested *AluI* fragments gave overlapping data for all the *AluI* sites in the DNA and yielded the complete sequence.

While additional primings were needed to confirm parts of the sequence, this approach can be seen to generate rapidly an overall sequence for the DNA. One can be sure that no fragments are missing, since all the overlaps are revealed and both ends of the sequence are known and the data are detailed enough to make a precise assessment of which additional restriction fragments are required to complete or confirm the sequence. In the random-cloning strategy the occurrence of identical or nearly identical fragments from a repeated region can sometimes lead to difficulty in identifying the presence of such a region. Eukaryotic DNA in

particular abounds in repeats and inverted repeats and a further feature of the approach described is that such regions are detected early in the sequencing strategy and their extent and position defined. Primers derived from such regions can hybridize to more than one site on the template and therefore give a mixture of different primed products. When a similar pattern of mixed sequences is also obtained by using the same primer on the complementary template the presence of a region of repeated sequence is confirmed. Provided that this region is not too extensive, primers which anneal to unique sequences flanking the repeated region may be used to sequence right through the repeat (Fig. 1). In Fig. 2 the sequence from position 228 to 575 shows the presence of $5\frac{1}{2}$ almost identical direct repeats of 62–63 nucleotides. The *AluI* fragment stretching from residues 314 to 439 is situated entirely within this repeat and all attempts to use this fragment as primer give mixed sequences due to priming at multiple sites.

The *Sau3A* sequence derived above was found to be identical, apart from two nucleotides, to the sequence of residues 2730–4309 in the yeast 2μ plasmid (Hartley & Donelson, 1980). The only difference between the two sequences is that whereas the 2μ plasmid contains the sequence ³⁶⁵⁷ATAGTCTTAC, this is replaced in pDAMY14 by ⁹²⁸ATAGCTTAC. This modification results in the introduction of a new *AluI* cleavage site, which was also confirmed in the present analysis. This near

GATCGTACTT GTTACCCATC ATTGAATTTT GAACATCCGA ACCTGGGAGT TTTCCCTGAA ACAGATAGTA TAATTGAACC TGTATAATAA TATATAGTCT 100
 AGCGCTTTAC GGAAGACATT GTATGTATTT CGGTTCCCTGG AGAACTATT GCATCTATTG CATAGGTAAT CTTCGACGTC GCATCCCGG TTCAATTTCT 200
 GCGTTTCCAT CTTCGACTTC AATAGCATAT CTTTGTGTTAAC GAAGCATCTG TGCCTTCAATT TGTAGAACA AAATGCAACG CGAGAGCGCT AATTTTTTCAA 300
 ACAAGAATC TGAGCTGCAT TTTTACAGAA CAGAAATGCA ACGCGAAAGC GCTATTTTAC CAACGAGAA TCTGTGCTTC ATTTTGTGTA AAAAAAATG 400
 CAACGCGAGA GCGCTAATTT TTCAAAACAA GAATCTGAGC TGCATTTTTA CAGAACAGAA ATGCAACGCG AGACGCTAT TTTACCAACA AAGAATCTAT 500
 ACTTCTTTTT TGTCTACAA AAATGCATCC CGAGAGCGCT ATTTTCTTAA CAAAGCATCT TAGATTACTT TTTTCTCTCT TTGTGCGCTC TATAATGCAG 600
 TCCTTTGATA ACTTTTTGCA CTGTAGGTCC GTTAAAGGTTA GAAGAAGGCT ACTTTGGTGT CTAATTTCTC TTCCATATAA AAAGCCTGAC TCCACTTTCCC 700
 GCGTTTACTG ATTACTAGCG AAGCTGCGGG TGCATTTTTT CAAGATAAAG GCATCCCGA TTATATTTCTA TACCGATGTG GATTGCGCAT ACITTTGTGAA 800
 CAGAAAGTGA TAGCGTTGAT GATCTTTTAT TGGTCAGAAA ATTATGAACG GTTCTTTCTA TTTTGTCTCT ATATACTACG TATAGGAAAT GTTTACATTT 900
 TCGTATTGTT TTCGATTAC TCTATGAATA GCTTACTACA ATTTTCTTGT CTAAAGAGTA ATACTAGAGA TAAACATATA AAATGTAGAG GTCGAGTTTA 1000
 GATGCAAGTT CAAGGAGCGA AAGGTGGATG GGTAGGTTAT ATAGGGATAT AGCACAGAGA TATATAGCAA AGAGATACIT TTGAGCAATG TTTGTGGAAG 1100
 CCGTATTCCG AATATTTTAT TAGCTCGTTA CAGTCCGGTG CGTTTTTGGT TTTTGTGAAAG TGCCTTTCTA GAGCGCTTTT GGTTTTCAA AGCGCTCTGA 1200
 AGTTTCTATA CTTTCTAGAG AATAGGAACT TCGGAATAGG AACTTCAAAG CGTTTCCGAA AACGAGCGCT TCCGAAATG CAACGCGAGC TGCGGCACATA 1300
 CAGCTCACTG TTCACGCTGC ACCATATATCT GCGTGTGTC TGTATATATA TATACATGAG AAGAACGGCA TAGTGGGTGT TTATGCTTAA ATGCGTACTT 1400
 ATATGCGTCT ATTTATGATG GATGAAAGGT AGTCTAGTAC CTCCTGTGAT ATTATCCCAT TCCATGCGGG GTATCGGTATG CTTCTTTCAG CACTACCTTT 1500
 TAGCTGTTCT ATATGCTGCC ACTCCTCAAT TGGATTAGTC TCATCCTTCA ATGCTATCAT TTCTTTTIGAT ATTGGATC 1578

Fig. 2. Sequence of the yeast replication origin-containing *Sau3A* fragment cloned in *pDAMY14*.

The sequence is identical, apart from two nucleotides, with a sequence within the yeast 2 μ plasmid (residues 2730–4309). The overlined sequences (residues 228–575) show the block of 5½ direct repeats and IR indicates the start of one of the inverted repeat sequences.

identity strongly suggests that the replication origin cloned in pDAMY14 did in fact arise from 2μ circle DNA in the original yeast DNA preparation and confirms the earlier suggestion (Hartley & Donelson, 1980) that this region of the plasmid sequence contains the origin of replication. This is further supported by evidence that yeast chromosomal DNA has no sequence homology with the 2μ plasmid (Cameron *et al.*, 1977). The slight difference between the two sequences (two nucleotides out of 1578) could be due to the different yeast strains used.

We thank Dr. David Beach for the gift of pDAMY14 and the Medical Research Council for support.

References

- Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282**, 189–194
- Beach, D., Piper, M. & Shall, S. (1980) *Nature (London)* **284**, 185–187
- Beggs, J. D. (1978) *Nature (London)* **275**, 104–109
- Cameron, J. R., Phillipsen, P. & Davis, R. W. (1977) *Nucleic Acids Res.* **4**, 1429–1448
- Clewell, D. B. & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. U.S.A.* **62**, 1159–1166
- Gronenborn, B. & Messing, J. (1978) *Nature (London)* **272**, 275–277
- Hartley, J. L. & Donelson, J. E. (1980) *Nature (London)* **286**, 860–864
- Heidecker, G., Messing, J. & Gronenborn, B. (1980) *Gene* **10**, 68–73
- Hollenberg, C. P., Degelmann, A., Kustermann-Kuhn, B. & Roger, H. D. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2072–2076
- Konings, R. N. H. (1980) *Methods Enzymol.* **65**(1), 795–811
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**(1), 499–560
- Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Schreier, P. H. & Cortese, R. (1979) *J. Mol. Biol.* **129**, 169–172
- Smith, A. J. H. (1979) *Nucleic Acids Res.* **6**, 831–848
- Yang, R. C.-A., Lis, J. & Wu, R. (1980) *Methods Enzymol.* **68**, 176–182