Sequence of 863 nucleotides encompassing the PstI site of the yeast 2 μ m plasmid

T.C.Elleman and J.Hindley

Department of Biochemistry, University of Bristol, Medical School, University Walk, Bristol, BS8 1TD, UK

Received 26 September 1980

ABSTRACT

The sequence of part of the larger unique region of the yeast 2 μ m plasmid cloned in pMB9 has been determined. The sequence extends from the single EcoRI site in this region to the AvaI site and includes the single PstI site and HpaI site. A notable feature of this sequence is the presence of tandem repeats of 124 residues beginning at the HpaI site and extending beyond the AvaI site. The sequence was determined independently by both the Maxam-Gilbert procedure applied to isolated restriction fragments, and by the chain-termination procedure applied to restriction fragments cloned in the single-stranded phage M13mp2 and purified by plaque selection.

INTRODUCTION

Certain strains of Saccharomyces cerevisiae (yeast) contain in the cytoplasm closed-circular, double-stranded DNA molecules of two microns circumference (1,2). This 2 μ m plasmid (Scp1) provides a vector for the cloning and expression of eukaryote genes in yeast as demonstrated by the expression of the Leu⁺ phenotype in a Leu⁻ yeast strain following transformation with the 2 μ m plasmid carrying an insertion from wild type S.cerevisiae chromosomal DNA in the PstI site (3).

The 2 μ m plasmid contains two 620 base-pair, non-tandem inverted repeat sequences separated by a larger (2.7 kbp) and a smaller (2.24 kbp) unique region (2,4,5,6). In yeast cells intramolecular recombination can occur within the inverted repeat sequences resulting in two populations of plasmid differing in the orientation of one unique region with respect to the other (4,7). Beggs (3) has cloned the yeast 2 μ m plasmid into the E.Coli plasmid pMB9 at the single EcoRI site of the latter. Eight full-size recombinant plasmids are theoretically possible depending on the yeast plasmid configuration (type A or B of (3)), which EcoRI site was opened, and the orientation of insertion into pMB9. In the present work we have used a type A recombinant pJDB71 (XY'form of (6)) which has the structure shown in Fig. 1.

Fig. 1: A diagrammatic representation of the recombinant plasmid pJDB71, Type A. The disposition of restriction enzyme targets is shown; R1 and R2 are *EcoRI* sites, H1, H2 and H3 are *Hind III* sites. The solid line represents the yeast 2 μ m plasmid linearized at site R2 by partial *EcoRI* digestion, and cloned into pMB9 represented by the dotted line. The disposition of the inverted-repeats is shown by the zig-zag portions.

The region of sequence studied is of interest since (i) it contains the single *PstI* site used for cloning (3) (ii) it contains the single *HpaI* site marking the beginning of a region distal to the *PstI* site which is absent in the shorter plasmids Scp2 and Scp3 with deletions of 0.13 and 0.22 kbp respectively (6). (iii) the mapping data of Broach et al. (8) indicate that this region contains one of the two major origins of transcription. Also several transcriptional products appear to terminate in this region. (iv) the replication origin has been mapped within, or very close to, the sequence determined (Broach, unpublished results).

EXPERIMENTAL PROCEDURES

The preparation and isolation of the recombinant plasmid pJDB71, the isolation of restriction fragments and their 5'-end labelling have been described previously (9).

Sequencing procedures

(i) The Maxam-Gilbert method.

The sequences of restriction fragments labelled at a single 5'-end were determined using the protocol described by Maxam and Gilbert (10) except that the amount of RNA carrier was reduced ten fold and the final denaturation step was carried out in formamide. Electrophoresis was carried out on thin sequencing gels as described by Sanger and Coulson (11).

(ii) By primed synthesis from restriction fragments cloned in phage M13mp2.

Restriction fragments were cloned into the EcoRI site of the double stranded form (RF1) of the single stranded bacteriophage M13mp2 using the double stranded EcoRI linker, 5'OH-GGAATTC-OH3', 3'OH-CCTTAAGG-OH5'. Following transfection into competent E.coli strain 71/18 (12,13) the mixture was plated out in the presence of the β -galactosidase inducer, IPTG and the chromogenic substrate 5-bromo-4-chloro-indolyl- β -galactoside (4). Colourless plaques were picked, transferred individually into 5 ml. medium and grown with shaking for 8-12 hours at 37°C. Recombinant phage was isolated from the cell free supernatant by precipitation with polyethylene-glycol 6000 and NaCl and single stranded phage DNA prepared by phenol extraction. Fig. 2

Fig. 3: Autoradiographs of sequencing gels prepared using the chain-termination method. Single-stranded template was from hybrid phage derived from M13mp2, by the insertion into the EcoRI site of AluI fragments spanning residues -5 to 75 and 129 to 76 respectively of the 2 μ plasmid. The M13mp2 sequence at the site of insertion is underlined.

DNA sequences were determined using the dideoxynucleotide chain termination procedure of Sanger, Nicklen and Coulson (15) as developed for sequencing M13mp2 cloned sequences (16,17). Fig. 3 shows sequencing gels obtained by this procedure from a cloned AluI digest of the 3.8 kbp EcoRI fragment.

RESULTS

The plasmid pJDB71 (type A recombinant) has the structure shown in Fig. 1. Digestion with EcoRI yielded three fragments which were well separated on a 1% agarose gel. The large, slowest-moving fragment represented pMB9 linearized at its single EcoRI site. Of the two other fragments (3.8 kbp and 2.4 kbp) derived from plasmid Scpl, the larger fragment containing the single PstI and HpaI sites was taken for further study. This fragment was digested with either Hae(II+III), AluI, or HpaI and the smaller fragments were cloned in M13mp2 after the addition of EcoRI linkers. The sequence obtained permitted the construction of one terminal sequence of the large 3.8 kbp fragment as shown in Fig. 4(a). The sequence is shown in Fig. 5. The bulk of this sequence had been previously ascertained by the isolation of end-labelled restriction fragments, the sequences of which were determined by the Maxam-Gilbert procedure (Fig. 4(b)). Both techniques showed absolute agreement in the sequences determined.

Fig. 4: Restriction endonuclease cleavages used to derive the PstI containing terminal region of the 3.8 kbp EcoRI cleavage fragment from yeast 2 μ m plasmid. -- represents the 5'-terminus of the single-stranded fragments used for sequence studies. The number of nucleotide residues between the PstI cleavage site and other cleavage sites are for the polarity of chain indicated. HhaI recognises the central tetranucleotide sequence of the HaeII recognition sequence, and the HpaI recognition sequence is one of the sequences recognised by HindII.

DISCUSSION

The combination of the chain-termination sequencing procedure of Sanger, Nicklen and Coulson with cloning in phage M13mp2 has greatly increased the speed with which DNA sequences can be determined. A complete digest is cloned in M13mp2 in a single experiment, thus involving no more extensive manipulation than is required for a single fragment. Ligation of a mixture of DNA fragments with a suitable excess of vector results in the insertion of single fragments into the vector. The recombinants are then separated merely by clone selection at an appropriate dilution on indicator plates.

We have noticed in our studies on unfractionated digests a dearth of clones containing larger fragments. This has necessitated some size fractionation of digests on gels so that smaller fragments which appear to clone at much higher frequencies are removed from the restriction digests prior to cloning.

The sequence of 863 nucleotides presented in Fig. 5, can be regarded as definitive since all sequences were determined at least twice in separate experiments and with the exception of residues -519 to -549 the sequences of both complementary strands were determined. The sequence beyond the AvaI site included in Fig. 6 relies on a single set of experiments but has been included because it is of special interest regarding the presence of repetitive sequences.

Beginning at or just before the HpaI site at position -314 a sequence of 124 residues is present (-315 to -438) which is repeated almost exactly (-439 to -563), followed by a further partial repeat. Two yeast plasmids Scp2 and Scp3 which show deletions of 0.13 kbp and 0.22 kbp in this region have been described (6). Both these plasmids lack the HpaI site, the deletion occurring on the PstI distal side of this site in Scp2, and probably similarly located in Scp3. Such deletions are easily accommodated in this region without the loss of any unique sequence since the absence of a complete repeat (125 kbp), or a complete plus partial repeat (210-220) kbp depending where the limits of this region are defined) only excludes redundant sequences. This repeated unit itself seems to be further constructed of two very similar halves of about 62 residues each (Fig. 6). No other extensive repeats or local regions of symmetry were found other than might be expected by chance.

One promoter proposed by Broach *et al.* (8) is thought to occur within the sequence shown in Fig. 5, and is probably located within the repetitive region. A hypothetical promoter site for interaction with RNA polymerase has

Fig. 5: The nucleotide sequence from the PstI-containing terminal region of the 3.8 kbp EcoRI fragment of yeast 2 μ m plasmid. Nucleotide residues are numbered from the PstI cleavage point in the chain shown. Local regions of symmetry are indicated: →, direct repeats shown above the sequence; + --- +, dyad symmetry shown below the sequence; [] palindromes.

```

      5'..ACGTCGCATCCCCGGTTCATTTTCTGCGTTTCCATCTTGCACCTTCAATAGC ATATCTTT
      HpaI
      ↓
-312  GTTAACGAAGCATCTGTGCTTCATTTT GTAGAACAAAAATGCAACGCGAGAGCGCTAATTTT

-374  TCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTT

-436  ACCAACGAAGAATCTGTGCTTCATTTTTGTAAACAAAAATGCAACGCGAGAGCGCTAATTTT

-499  TCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTT

-561  ACCAACAAAGAATCTATACTTCTTTTTTGTCTACAAAAATGCATCCGAGAGCGCTATTTT
                                   ↑
                                   AvaI

-623  TCTAACAAAGCATCTTAGCTACTTTTT..3'

```

Fig. 6: The sequence of residues -253 to -649 of the yeast 2 μ m plasmid, arranged to maximise sequence homology within this region. The sequence shows the complementary chain to that region shown in Fig. 5 although base-pair numbering is preserved.

been proposed for eukaryotic genes - the Hogness box - on the basis of homologous sequences upstream of several startpoints of transcription (19,20). Sequences similar to the 'Hogness box', are found in the repetitive region, and might represent a possible promoter site. Thus, the sequence GGTAAATAGCGC comprising residues -563 to -551 (Fig. 5) bears marked homology with several promoters. The site is followed 'downstream' by a region of partial dyad symmetry centred at residue -530. Such symmetry regions are often found around origins of transcription where they interact with regulatory proteins controlling transcription. Because of the low precision with which transcripts have so far been localized on the 2 μ m plasmid, similar functions might be postulated for the homologous sequences found 62 residues 'upstream' or 'downstream' of the example cited.

The pentanucleotide sequence AATAA is found near the 3' end of many eukaryotic messengers (21). This sequence occurs in both complementary strands of the region studied. The sequence at position -163 to -167 is well situated to correspond with the 3' end of the counterclockwise transcript originating from the second promoter and ending in this region (8), but the sequence at position 15 to 19 is probably not related to the 3' end of the clockwise transcript which is believed to terminate near the EcoRI site, outside the sequence studied. The localization of RNA polymerase recognition sites must await more precise mapping data on origins and terminations of

transcription.

Since this work was completed Hartley and Donelson (22) reported the entire sequence of the type A yeast plasmid recovered from S.cerevisiae strain A364 A D5 by cloning into pMB9. In our work the 2 μ plasmid originated from the unrelated strain S.cerevisiae DR19/8T (3) cloned into pMB9. Comparison of the two sets of sequence data show complete agreement and allow two conclusions to be drawn. Firstly, the results completely validate the reliability of the newer sequencing procedures used in this work. Neither during propagation of the hybrid plasmid nor in the subsequent M13 cloning and primed syntheses is there any evidence of rearrangements or deletions of sequences. Secondly, the data implies that at least one unique region in the plasmid, bounded by the inverted repeats, is fully conserved in two unrelated yeast strains.

Though the function of the yeast plasmid is unknown the lack of drift found in the sequence argues for its having some definite role in the organism rather than having a purely 'selfish' role where its only function is self replication (23).

ACKNOWLEDGEMENTS

We are grateful to Dr. J.D. Beggs for the cloned yeast plasmid, Dr. R. Cortese for the phage M13mp2am4, and Drs. J. Messing and B. Groneborn for their protocols prior to publication. We are much indebted to Dr. G. Winter for his advice of many experimental procedures used. J.H. is grateful to the Medical Research Council, U.K. for financial support, and T.C.E. to CSIRO, Australia.

REFERENCES

1. Guerineau, M., Slonimski, P.P. and Avner, P. (1974) Biochem. Biophys. Res. Commun. 61, 462-469.
2. Guerineau, M., Grandchamp, C. and Slonimski, P.P. (1976) Proc. Natl. Acad. Sci. USA 73, 3030-3034.
3. Beggs, J.D. (1978) Nature 275, 104-109.
4. Hollenberg, C.P., Degelmann, A., Kustermann-Kuhn, B. and Roger, H.D. (1976) Proc. Natl. Acad. Sci. USA, 73, 2072-2076.
5. Livingston, D.M. and Klein, H.L. (1977) J. Bact. 129, 472-481.
6. Cameron, J.R., Phillippsen, P. and Davis, R.W. (1977) Nucleic Acids Res. 4, 1429-1448.
7. Beggs, J.D., Guerineau, M. and Atkins, J.F. (1976) Mol. Gen. Genet. 148, 287-294.
8. Broach, J.R., Atkins, J.F., McGill, C. and Chow, L. (1979) Cell, 16, 827-839.
9. Hindley, J., and Phear, G.A. (1979) Nucleic Acids Res. 7, 361-375.

10. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 560-564.
11. Sanger, F. and Coulson, A.R. (1978) *FEBS Letters*, 87, 107-110.
12. Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 3642-3646.
13. Gronenborn, B. and Messing, J. (1978) *Nature* 272, 275-377.
14. Miller, J. (1972) in *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463-5467.
16. Schreier, P.H. and Cortese, R. (1979) *J. Mol. Biol.* 129, 169-172.
17. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. *J. Mol. Biol.* (1980) in the press.
18. Zain, B.S. and Roberts, R.J. (1979) *J. Mol. Biol.* 131, 341-352.
19. Proudfoot, N.J. (1979) *Nature* 279, 376.
20. Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garaphin, A., Cami, B., Chambon, P. (1979) *Nature* 278, 428-434.
21. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
22. Hartley, J.L. and Donelson, J.E. (1980) *Nature* 286, 860-864.
23. Orgel, L.E. and Crick, F.H.C. (1980) *Nature* 284, 604-607.