Volume 9 Number 2 1981 Nucleic Acids Research

# A system for shotgun DNA sequencing

Joachim Messing[+†], Roberto Crea[*] and Peter H.Seeburg[**]

[+]Department of Bacteriology, UC Davis, CA 95616, and [*]Division of Organic Chemistry and [**]Division of Molecular Biology of Genentech, Inc., South San Francisco, CA 94080, USA

Received 30 October 1980

ABSTRACT

A multipurpose cloning site has been introduced into the gene for β-galactosidase (β-D-galactosidegalactohydrolase, EC 3.21.23) on the single-stranded DNA phage M13mp2 (Gronenborn, B. and Messing, J., (1978) Nature 272, 375-377) with the use of synthetic DNA. The site contributes 14 additional codons and does not affect the ability of the lac gene product to undergo intracistronic complementation. Two restriction endonuclease cleavage sites in the viral gene II were removed by single base-pair mutations. Using the new phage M13mp7, DNA fragments generated by cleavage with a variety of different restriction endonucleases can be cloned directly. The nucleotide sequences of the cloned DNAs can be determined rapidly by DNA synthesis using chain terminators and a synthetic oligonucleotide primer complementary to 15 bases preceeding the new array of restriction sites.

## INTRODUCTION

Understanding of genes from higher organisms requires the determination of long stretches of nucleotide sequences. Since currently used methods for DNA sequencing (1,2) are relatively time-consuming, a rapid sequencing system is presented involving a shotgun approach. The system allows for easy cloning of short overlapping DNA fragments form and their rapid sequence determination by primed DNA synthesis (2) using only one primer preparation for all templates.

Certain single-stranded DNA phages (e.g. M13) are packaged and extruded by their hosts even if foreign DNA has been inserted into their circular genomes (3-8). This feature is especially attractive for a fast preparation of single-stranded template from any DNA cloned into the phage RF-DNA. To facilitate cloning, part of the lac operon from E. coli had been inserted into M13 RF-DNA and an EcoRI restriction site created in a region coding for the N-terminus of β-galactosidase (9,10). The lac region in this phage M13mp2 serves as a marker system for distinguishing vector DNA (blue plaque-formers) from recombinant DNA (white

plaque-formers). Cloned DNA can be sequenced with the help of a primer, itself constructed from lac DNA (11) and cloned into the single EcoRI site in plasmid pBR325 (12).

Although any DNA fragment can be inserted into the EcoRI site of M13mp2 RF-DNA by attaching the appropriate sticky ends using oligonucleotide linkers (13), a different approach is described here to convert M13 into a system for shotgun DNA sequencing. With the help of synthetic DNA an array of endonuclease restriction sites has been introduced into the lac region of the phage. This allows direct cloning of DNA fragments generated by a number of different restriction enzymes which yield cohesive or blunt-ended termini. The successful insertion of such DNA fragments can still be monitored by the color of the plaques (9). To facilitate large-scale sequencing of recombinant templates a single-stranded oligonucleotide of 15 residues was chemically synthesized to replace the biological primer (11).

## Materials and Methods

Phages M13mp2, M13mp6, and bacterial strains JM101, K38, K37 are described elsewhere (9,14,15). E. coli JM103 was derived from JM101 and SK1592 (16) and has the following genotype: Δlacpro, thi, strA, supE, endA, sbcB15, hsdR4, F'traD36, proAB, $lacI^q$ ZΔM15. Phages MWJ43 (17) and M13 am1, am2 were kindly provided by Ray Wu and David Pratt respectively. Phages M13mp61, M13mp62, M13mp63, and M13mp71 are intermediates in the construction of M13mp7. M13mp61, like its parent M13mp6, has resistance to endonuclease BamHI in gene III (14) but contains in addition 2 amber mutations. M13mp62 and M13mp63 carry additional resistances in gene II to HincII, and to HincII and AccI respectively. M13mp71 contains the synthetic 17-mer specifying SalI and PstI restriction sites ligated into MWJ43 RF-DNA. M13mp7 carries two amber mutations and is resistant to BamHI, HincII and AccI in the M13 part of its genome.

Single-stranded and double-stranded DNA of M13 phage was prepared as described (3,11). Restriction endonucleases were obtained from New England Biolabs and used as described by the supplier. $T_4$ DNA ligase was isolated from a lambda lysogenic strain (18). Conditions for a sticky end or blunt end ligation were the same as indicated in the isolation procedure (19). Polyacrylamide gel electrophoresis, ethidium bromide staining of DNA, and electroelution of DNA from polyacrylamide

gels and DNA sequencing with chain terminators were as described (2,11). Synthesis of oligonucleotides was conducted using a modified triester method (20). The 17mer, 5'-GATCCGTCGACCTGCAG-3', was used as adapter for enzymes SalI and PstI into a BamHI site. The 15mer, 5'-CCCAGTCACGACGTT-3' is complementary to the lac region of the phage DNA 17 nucleotides before the EcoRI cloning site.

To convert M13mp6 into M13mp61 the RF-DNA of M13 am1, am2(1µg) was cleaved with BamHI and HaeIII, and fragments denatured (100°C, 2 min) and renatured (66°C, 3 hours) in the presence of M13mp6 viral DNA (50ng). The resulting heteroduplex with two internal mismatches, one gap of the length of lac DNA and one 3' terminal mismatch corresponding to the 5'G of the BamHI-cleaved RF-DNA was used to transform JM103. Single plaque formers were transferred to soft agar layers of K38 ($su^-$) and K37 ($su^+$). Phages not multiplying on K38 lawns were saved on a K37 replica plate and reversion rates were tested. About 24 percent of all the transformants did not grow on K38 and 80 percent of these were double ambers. No colorless plaques were observed from a total of 230 transformants. M13mp61 produces normal phage titers ($5x10^{12}$/ml).

Marker rescue with a double amber was performed as follows. RF-DNA(1µg) from the double amber was cleaved with the restriction endonuclease as indicated, and DNA fragments were purified by phenol treatment, precipitated with ethanol and redissolved in 20 µl of 50 mM Tris-HCl, pH 7.6, 100 mM NaCl, and 1mM EDTA. A viral DNA solution (1µl, 50ng) containing the (+) strand for hybridization was added to the 20 µl of DNA fragments containing the (-) strands. The mixture was held for 3 minutes in boiling water, and further incubated at 66°C for 3 hours. Incubation was terminated by putting the mixture on ice. Host cells were made competent as described by Cohen et al. (21). Plating conditions were as described previously (3). A few µl of the hybridization mix gave rise to approximately 300 transformants.

Base specific mutations of C→T in endonuclease HincII and AccI recognition sites were performed with hydroxylamine treatment of the particular phages as described (15). After hydroxylamine treatment, the initial phage titer dropped from $3x10^{12}$ to $2x10^6$. The mutagen was removed after 3 hours by infecting host cells and washing them twice in medium. Surviving phages were grown and RF-DNA isolated. The selection was initiated by treating RF-DNA twice with HincII or AccI endonuclease respectively to assure cleavage of RF-DNA lacking the particular

mutation. About 5 µg of cleaved RF-DNA was used to transform JM103. An aliquot of uncleaved RF-DNA served as a control in a separate transformation experiment. Comparing the respective transformation efficiencies the mutant was enriched 2,500-fold. An approximate number of 2000 plaque formers was obtained and RF-DNA prepared from them was cleaved again for a second enrichment (100-fold). A single transformant was checked for its resistance to the endonuclease used in the selection.

All recombinant phages have been grown in JM101 or JM103 under P1 conditions according to the NIH guidelines (14) since these hosts contain the traD36 mutation (22).

## RESULTS

### Construction plan for a multipurpose cloning site

Chemical synthesis of DNA as well as enzymatic reactions involving DNA ligase and polymerase have been used to contruct a multipurpose cloning site in the lac region of a phage M13 derivative. The procedures are outlined in Fig. 1. The adapter used to insert SalI and PstI sites was a 17mer with a 5'-terminal sequence complementary to the cohesive ends of the BamHI site in the M13mp2 BamHI phage MWJ43 (17). The final product (M13mp71) has 24 nucleotides more than the BamHI phage, thus conserving the reading frame of the gene. The total number of codons additional to those in the lac region of M13mp2 is 14.

### Synthesis and insertion of an adapter for SalI and PstI

The adapter for the introduction of SalI and PstI sites was the chemically synthesized oligonucleotide 5'-GATCCGTCGACCTGCAG-3'. To insert the adapter into phage MWJ43 it was labeled at its 5' end using [$\gamma$-$^{32}$P]ATP and T4 polynucleotide kinase, and ligated to the BamHI-cleaved RF-DNA. The ligation products were separated by polyacrylamide gel electrophoresis, and the gel examined by ethidium bromide straining as well as autoradiography (Fig. 2). The DNA visible as a dark band on X-ray film was recovered from the gel to remove unincorporated adapter and recircularized parental molecules. The DNA contained 3'-protruding ends of which the last 6 nucleotides were self-complementary (see Fig. 1). To stabilize the corresponding circular form this DNA was hybridized to itself and gaps were filled using reverse transcriptase in the presence of deoxynucleotides. The filled-in DNA was

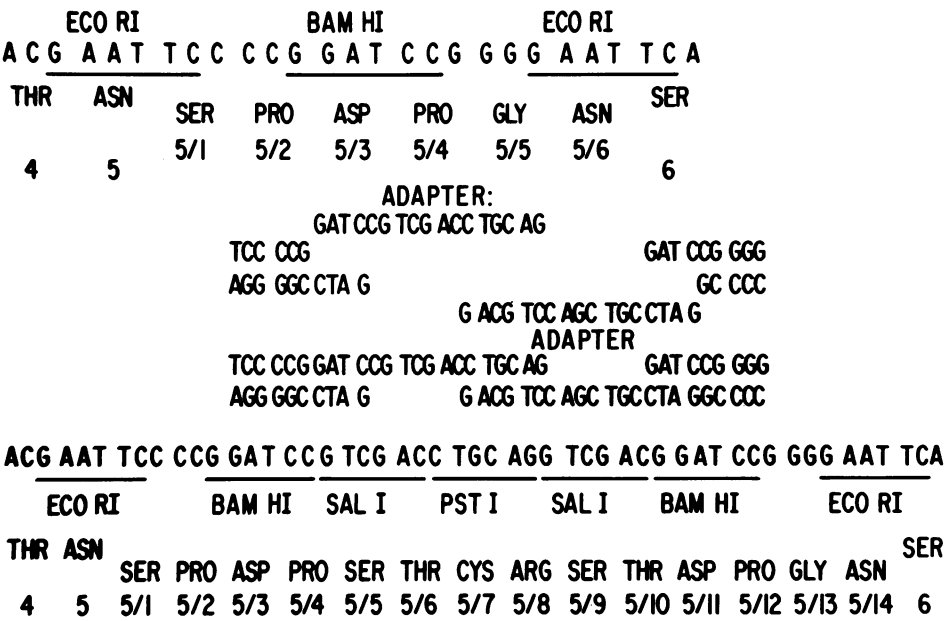



Fig. 1. Schematic outline of the construction of M13mp71
Shown are the nucleotide and amino acid sequences involved in the construction of a lacZ gene with 14 additional codons. Nucleotide sequences are given in the polarity of the (+) or messenger strand. Restriction sites are underlined. The numbers 4, 5, 6 refer to the positions of amino acids in β-galactosidase as encoded by M13mp2 (9). Dashed numbers indicate inserted codons. The upper section shows the cloning sites in phage MWJ43 RF-DNA (17). The middle section depicts the alignment of the adapter sequence of 17 nucleotides with the BamHI ends of phage MWJ43 RF-DNA, and the annealing of the PstI sticky ends to allow gap filling. The lower section shows the primary structure of the new lac region in M13mp71.

used to transform competent cells of E. coli JM101 (14) and 10 blue plaques were obtained on a bacterial lawn. No plaques formed if the filling reaction was omitted. Recombinant phage from one blue plaque was grown and the RF-DNA analyzed. As predicted, the DNA could be cleaved by endonucleases SalI and PstI. Although in this phage (M13mp7[illegible]) only the PstI site is unique, sites for SalI, BamHI and EcoRI are also usable for cloning because the ensuing loss of small inserts still results in a functional lac sequence (see Fig. 1).

It is of interest to note that the correct insertion of the synthetic 17-mer resulted in a recombinant phage which still forms blue plaques.

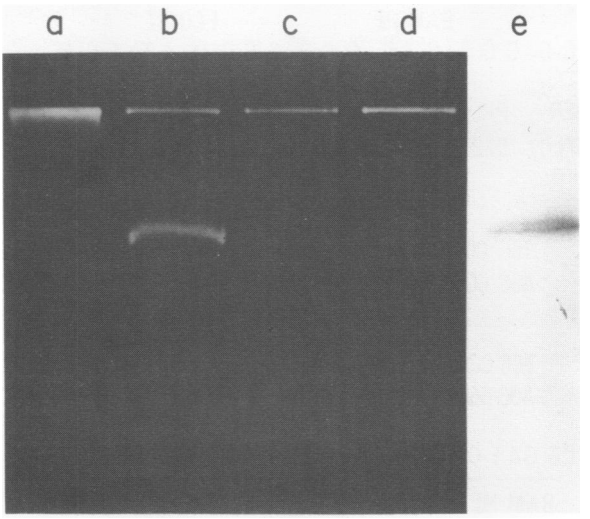



Fig. 2 Analysis of the ligation product between 17mer and phage DNA
MWJ43 RF-DNA was cleaved with BamHI and incubated with T4 DNA ligase at 8°C for 1 hour in the presence of a 100 fold molar excess of the 17mer radioactively labelled at its 5' end. The ligase was heat inactivated and the reaction mixture subjected to electrophoresis on a 5 polyacrylamide gel. DNA was visualized by ethidium bromide staining. Under the electrophoresis conditions used circular and high molecular weight DNA did not enter the gel (a). Only the linearized phage RF-DNA migrated a short distance into the gel (b). As a control the BamHI-cleaved DNA was ligated without the 17mer (c). Only the reaction including the 17mer showed ligation products, mainly in the position of linear phage DNA (d). When the gel was exposed to X-ray film only the reaction products in (d) in the position of linear DNA gave a radioactive signal (e).

Only the lesser intensity of the color distinguishes it from M13mp2. Similar shifts in intensity have been previously noted with phages containing one additional cloning site integrated into M13mp2 by insertion through EcoRI sticky ends of a DNA fragment several codons long (14,17).

## The versatility of the SalI Site

The sequence 5'-GTCGAC-3' is cleaved by endonuclease SalI to produce cohesive termini of 4 nucleotides which are complementary to those generated by endonuclease XhoI. The same sequence is recognized by endonuclease AccI which cleaves it to produce the same sticky ends (pCG-) as do enzymes ClaI, TaqI, and HpaII. Furthermore endonuclease HincII cleaves this sequence to generate blunt ends. Unfortunately, due to

ambiguities in their recognition sequences (23), the two endonucleases AccI and HincII each cleave one additional site, both located in gene II of M13 (24). To fully utilize the versatility of the SalI site in M13mp71 these additional cleavage sites were removed by chemical mutagenesis (Fig 3). To keep the phage viable mutagenesis was designed to change a nucleotide in the recognition site but not alter the encoded amino acid.

Regarding the HincII site 5-'GTTAAC-3' in M13 DNA gene II, only the C can be modified by hydroxylamine to mispair with an A. The subsequent substition of an AAC codon by an AAT codon is silent and does not change the gene II product (Fig. 3). The resulting HincII-resistant phage was

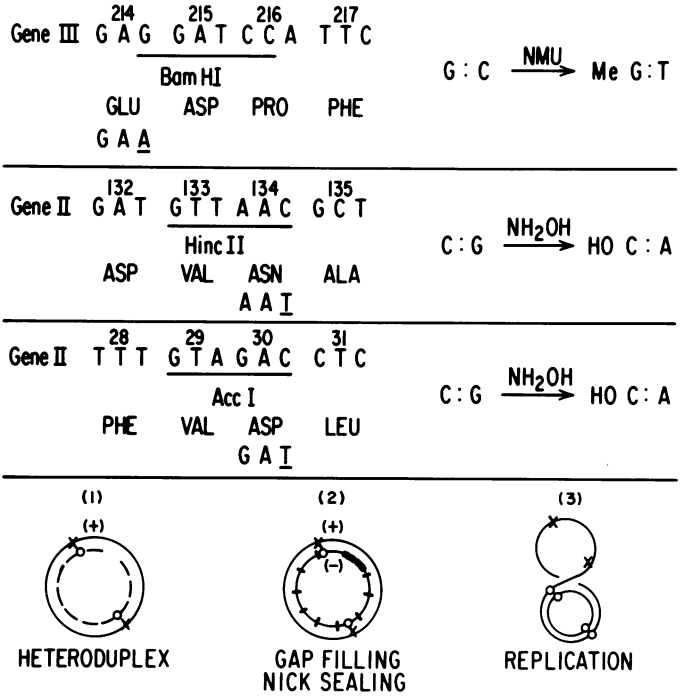


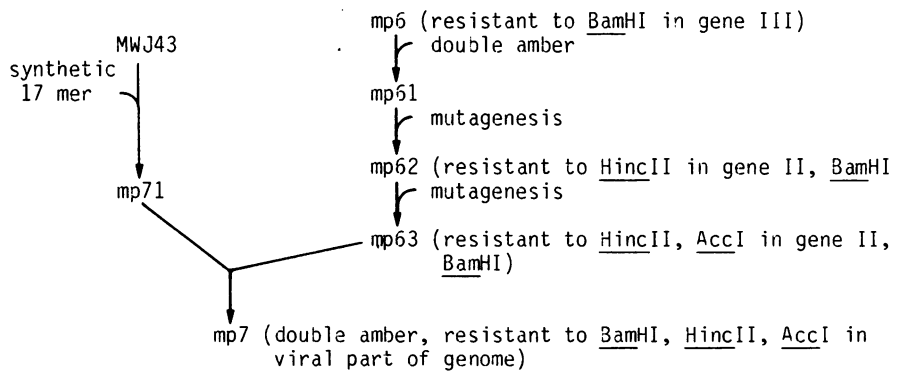

Fig. 3 Schematic representation of site directed in vitro mutagenesis
Nucleotide sequences are shown in 5'→3' direction of the viral strand. The presumed mechanism of mutagenesis is indicated for each case, G→A by methylation of the $O_6$ position of G, C→T by hydroxylamine. The two different codons for the same amino acid in the gene product are indicated. The transfer of such a mutation from the (-) strand to the (+) strand is represented by the flow of intermediate steps, going from a heteroduplex (1) to a complete parental RF in vivo (2) and finally to the replicative intermediate (3) in which the (-) strand is the template (25). The mutagenesis of the BamHI site in gene 3 (24) has been previously described (14) and is included for completeness.

mutagenized in the same way to substitute a GAC codon for a GAT codon, rendering the phage AccI-resistant (Fig. 3). This mutation creates a unique BglII site which can be used to demonstrate the specific base change.

Mutations in the restriction sites were selected by transforming host cells with RF-DNA, obtained from a population of mutagenized phage and cleaved with the appropriate endonuclease. Linearized RF-DNA will tranform host cells with a much-lower efficiency than circular RF-DNA (9). Since this type of selection requires the cleavage site to be unique, removal of the two sites from M13 had been performed on phage strains (M13mp61 and M13mp62) lacking the array of newly constructed cloning sites.

The proper mutations were subsequently transferred to M13mp71. To conveniently monitor this transfer two amber mutations had been introduced into phage M13mp6 resulting in phage M13mp61. The presence of these mutations in M13mp61 allowed a marker rescue type of experiment (25) in which differential plating efficiencies on $su^-$(K38) and $su^+$ (K37) hosts were used as an easy screen for the cotransfer of mutations in the restriction sites. Cotransfer of several mutations was performed via heteroduplex formation between restriction fragments of RF-DNA from mutant phage and intact viral DNA of the acceptor phage. Such cotransfer is possible because the minus strand is the template for replication in a rolling circle mechanism (26) and internal mismatches are not repaired.

The manipulations described above are illustrated in the following flow chart:




The end product is phage M13mp7 in which the SalI site can be also used

for cloning blunt-ended DNA after cleaving the RF-DNA with HincII, and pCG-sticky ends after AccI cleavage.

A synthetic primer for DNA sequencing with M13mp7

The several cloning sites in M13mp7 allow for direct insertion of DNA fragments generated by a variety of different restriction endonucleases. A detailed cloning guide is shown in Fig. 4. Since all cloning sites are located next to one another inserted DNA can be sequenced using only one primer preparation to serve all sites. Such a primer was chemically synthesized as an oligonucleotide of 15 residues. It has the sequence of 5'-CCCAGTCACGACGTT-3' and was designed to prime within the lac region 17 nucleotides from the array of cloning sites. The specificity of this primer is demonstrated in Fig. 5 The synthetic primer thus substitutes for the one previously used on M13mp2, first cloned as an EcoRI fragment of 90 base-pairs (11), and later reduced in size (27).

## DISCUSSION

The combination of the E. coli lac system and the filamentous single-stranded DNA phage M13 provides a biological system for preparing cloned templates in a single-stranded form required for DNA sequencing with chain terminators (2). There is a concern regarding the stability of the cloned DNA in recombinant M13 phage. Both phages, M13mp2 and M13mp5 have been used to clone DNA of 1000-3000 base pairs and to sequence it using different restriction fragments as primers (28-30). DNA sequences obtained in this way were identical to sequences obtained from plasmid clones using the method of Maxam and Gilbert (1). Although viral DNA up to 6 times the length of M13 DNA has been packaged (Messing, unpublished), deletions may occur within the cloned DNA (Messing, unpublished, 28). Recombinants with large inserts, however, can be plaque - purified and subjected to a minimal number of bursts to obtain the desirable amount of phage or RF-DNA.

Since deletions in the cloned DNA will be minimized if it is small, DNA of large size can be stored in subclones of smaller restriction fragments. The new phage M13mp7 is a versatile vehicle for preparing such fragment banks from any DNA. Moreover with a primer available, M13mp7 is an excellent vector for a shotgun approach to obtain rapidly the nucleotide sequence of any DNA of interest. The DNA of interest is converted into sets of fragments, each set generated by one of the

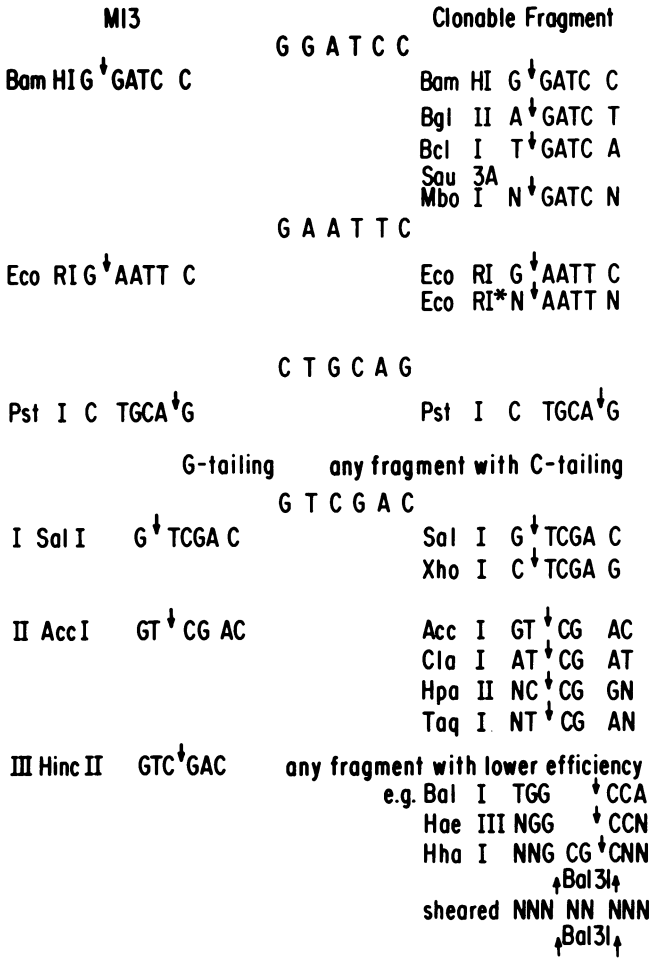



Fig. 4 Cloning guide for M13mp7

Since sticky ends are in many cases only a part of the restriction endonuclease recognition site, DNA cleaved at different sites with a common sticky end can be joined together. Although this leads to hybrid sites which are not recognized by either of the enzymes, the flanking restriction sites can be used to release off the fragment (see also Fig. 1). The use of a common set of sticky ends increases the versatility of cloning into M13mp7 RF DNA. Therefore the enzymes and their recognition sequences are used for cleaving the vector are listed on the left side. The enzymes and treatments used to produce DNA fragments which may be cloned at each M13mp7 site are listed on the right side. Cutting patterns indicated by arrows and sticky ends are set apart from flanking nucleotides.

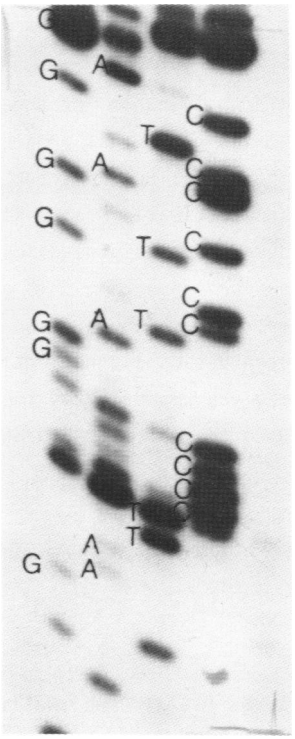



Fig. 5 Specificity of the synthetic primer
A recombinant phage containing a Pst I restriction fragment has been used as a template to demonstrate the specificity of the synthetic primer. Preparation of template, sequencing reaction, electrophoresis conditions are described (2,11). Reactions were loaded in the GATC order. A photograph of the autoradiogram is shown covering the multiple restriction sites in M13mp7 RF-DNA. The insert can be cleaved from the RF-DNA with PstI, SalI, BamHI or EcoRI (not shown).

different restriction endonucleases listed in Fig 4. When cloned these sets yield banks of recombinant M13mp7 phage in which the large DNA is represented piece-meal with the two complementary DNA strands present in separate clones. Since different fragment sets are overlapping with each other in sequence, the entire primary structure of the DNA can be reconstructed by compiling the data obtained from sequencing several recombinant M13mp7 banks and aligning the nucleotide sequences by overlaps and complementarities. The large amount of data processing necessary for reconstructing long sequences can be efficiently assisted

by computer (31).

Acknowledgements

We thank David Pratt and John Ingraham for their helpful discussion and for making facilities available for this project. We thank the Dean of Agriculture and Environmental Sciences for his continuous interest and his financial support, David Pratt, Ray Wu for strains and Richard Gardner, Gisela Heidecker for helpful discussions. We are grateful to Robert A. Swanson for supporting this project. The excellent technical help of Mark Vasser, Parkash Jhurani and Martin Struble in preparing and analyzing the synthetic oligonucleotides is gratefully acknowledged. Joachim Messing was supported by the Deutsche Forschungsgemeinschaft and by grant no. 5901-9-0386 of the United States Department of Agriculture.

†Present address: Department of Biochemistry, University of Minnesota, St Paul, MN 55108, USA

References

1. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
2. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
3. Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) Proc. Natl. Acad. Sci. USA 74, 3642-3646.
4. Barnes, W.M. (1978) Proc. Natl. Acad. Sci. USA 75, 4281-4285.
5. Ray, D.S. and Kook, K. (1978) Gene 4, 109-119.
6. Ohsumi, M., Vovis, G.F. and Zinder, N.D. (1978) Virology 89, 438-449.
7. Herrmann, R., Neugebauer, K., Zentgraf, H. and Schaller, H. (1978) Molec. Gen. Genet. 159, 171-178.
8. Nomura, N., Yamagishi, H. and Oka, A. (1978) Gene 3, 39-48.
9. Gronenborn, B. and Messing, J. (1978) Nature 272, 375-377.
10. Messing, J. and Gronenborn, B. (1978) In: Single-Stranded DNA Phages, eds Denhardt, D.T., Dressler, D.H. and Ray, D.S. (Cold Spring Harbor Laborat., Cold Spr. Harb., N.Y. 11724), pp. 449-453.
11. Heidecker, G., Messing, J. and Gronenborn, B. (1980) Gene 10, 69-73.
12. Bolivar, F. (1978) Gene 4, 121-136.
13. Scheller, R.H., Dickerson, R.E., Boyer, H.W., Riggs, A.D. and Itakura, K. (1977) Science 196, 177-180.
14. Messing, J. (1979) Recombinant DNA Technical Bulletin 2, 43-48.
15. Pratt, D., Tzagoloff, H. and Erdahl, W.S. (1966) Virology 30, 397-410.
16. Kushner, S.R. (1978) In: Genetic Engineering, eds Boyer, H.W. and Nicosia, S., (Elsevier/North-Holland Biomed. Press, Amsterdam, Netherl.) pp.17-23.

17. Rothstein, R.J., Lau, L.F., Bahl, C.P., Narang, S.A. and Wu, R. (1980) In: Methods in Enzymology, eds Colowick, S.P. and Kaplan, N.C., (Academic Press, Inc., New York), Vol 68, pp. 101-110.
18. Murray, N.E., Bruce, S.A. and Murray, K. (1979) J. Mol. Biol. 132, 493-505.
19. Tait, R.C., Rodriguez, R.L. and West, Jr., R.W. (1980) J.Biol.Chem. 255, 813-815.
20. Crea, R., Kraszewski, A., Hirose, T. and Itakura, K. (1978) Proc. Natl. Acad. Sci. USA 75, 5765-5769.
21. Cohen, S.N., Chang, A.C.Y. and Hsu, C.L. (1972) Proc. Natl. Acad. Sci. USA 69, 2110-2114.
22. Achtman, M., Willets, N. and Clark, A.J. (1971) J. Bacteriol. 106, 529-538.
23. Roberts, R. (1980) Nucl. Acids. Res. 8, 63-80.
24. Van Wezenbeck, P., Hulsebols, T. and Schoenmakers, J.G.G., (1980) Gene, 11, 129-148.
25. Hutchinson, C.A., Phillips, S., Eolgell, M.H., Gillam, S., Jahnke, P. and Smith, M. (1978) J. Biol. Chem. 253, 6551-6560.
26. Staudenbauer, W.L., Kessler-Liebscher, B.E., Schneck, P.K., van Dorp, B. and Hofschneider, P.H. (1978) In: Single-Stranded DNA Phages, eds Denhardt, D.T., Dressler, D.H. and Ray, D.S. (Cold Spr. Harb. Laborat., Cold Spring Harb., N.Y. 11724), pp. 369-378.
27. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucl. Acids. Res. 8, 1731-1743.
28. Cordell, B., Bell, G., Tischer, E., DeNoto, F.M., Ullrich., A., Pictet, R., Rutter, W.J. and Goodman, H.M. (1979) Cell 18, 533-543.
29. Buchel, D.E., Gronenborn, B. and Muller-Hill (1980) Nature 283, 541-545.
30. Czernilofsky, A.P., Delorbe, W., Swanstrom, R., Varmus, H.E., Bishop, J.M., Tischer, E., and Goodman, H.M, (1980) Nucl. Acids Res. 8, 2967-2984.
31. Staden, R. (1977) Nucl. Acids Res. 4, 4037-4051.
32. Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
33. Jay, E., Bambara, R., Padmanabhan, P. and Wu, R. (1974) Nucl. Acids Res. 1, 331-353.