
The nucleotide sequence of the *B* gene of bacteriophage Mu

Janet L. Miller¹, Stephen K. Anderson¹, Donald J. Fujita¹, George Chaconas^{1*}, Donald L. Baldwin² and Rasika M. Harshey²

¹Cancer Research Laboratory and Dept. Biochemistry and Microbiology and Immunology, University of Western Ontario, London, Ontario N6A 5B7, Canada, and ²Research Institute of Scripps Clinic, Dept. Molecular Biology, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

Received 31 August 1984; Revised and Accepted 22 October 1984

ABSTRACT

Bacteriophage Mu is a highly efficient transposon which requires the products of the Mu A and B genes in order to transpose at a normal frequency. We have determined the nucleotide sequence of the B gene as well as that of the A-B intergenic region upstream of B. The protein product of the gene contains 312 amino acids and has a predicted molecular weight of 35,061. As expected, there do not appear to be any potential promoter sequences in the intergenic region prior to the gene, but it is preceded by a strong Shine-Dalgarno sequence. The intergenic region does not contain any obvious transcription termination sequences. The frequency of optimal codon usage is similar to that for other transposon and phage genes, and the amino acid composition is comparable to that of an "average" *E. coli* protein. A region near the amino terminus of the protein resembles the highly conserved bihelical fold which is involved in DNA contact and sequence specific recognition in a number of DNA binding proteins.

INTRODUCTION

Bacteriophage Mu is remarkable for its ability to transpose at a frequency which is orders of magnitude higher than that for other transposons (see ref. 1 for a recent review). Transposition is an integral part of the Mu life cycle and can be either conservative or replicative in nature. The transposition pathway used appears to be governed by the source of the transposing DNA. Following infection with phage Mu, transposition of the parental DNA occurs to yield simple insertions (2) in what is a conservative integration event (3). In contrast, transposition of Mu DNA which is already resident in the host cell generates predominantly cointegrates, which are replicative by nature (4).

A number of host proteins are required for transposition (5), but the Mu A and B gene products are the only phage

proteins required for normal levels of transposition (6,7,8). The A protein is essential for both types of transposition and is the Mu encoded transposase (6,7,8). The B protein seems to be responsible for the high frequency of transposition which distinguishes Mu from other transposons. In the absence of the B protein replicative transposition still occurs, but at a frequency which is reduced about 100 fold (7,8,9). Mu DNA replication does not occur at a significant level in either A⁻ or B⁻ mutants (10,11). The B protein may also play a role in determining whether the end-products of transposition are simple insertions or cointegrates (9), but the data on this are as yet contradictory, and no clear conclusion can be drawn.

In order to aid in the determination of the role of the B protein in Mu DNA transposition, we sequenced the entire B gene, as well as the A-B intergenic region upstream of B. The B gene was found to code for a protein with a molecular weight of 35,061. The codon usage, amino acid composition of the protein, and the DNA regulatory sequences of the gene are discussed.

MATERIALS AND METHODS

Bacterial and phage strains

Bacteriophage M13mp8 and its host strain JM103 were obtained from Bethesda Research Laboratories (BRL). The M13 phages were propagated as described by the BRL manual.

Enzymes and reagents

Restriction enzymes were obtained from BRL and New England Biolabs, and the digestion conditions recommended in Davis et al. (12) were used. T4 DNA polymerase, T4 DNA ligase, bacterial alkaline phosphatase and Klenow polymerase were purchased from BRL. RNase T1 was supplied by Sankyo. T4 polynucleotide kinase was from New England Biolabs. The reagents for dideoxy sequencing were purchased from the following companies: dideoxy NTPs were from Collaborative Research, dNTPs were from Pharmacia or BRL, the 17 bp primer was from New England Biolabs (#1211) and the [α -³⁵S]dATP was from New England Nuclear. [α -³²P]dATP and [γ -³²P]dATP for Maxam and Gilbert sequencing were obtained from Amersham.

Plasmid DNA Preparation

Plasmid DNA was prepared by the method of Birnboim and Doly (13), phenol extracted once and then run through a Sephadex G-50 mini-column (14). If the plasmid DNA was to be restricted it was treated with RNase T1 at the same time.

Construction of M13 clones

M13 clones for DNA sequencing were constructed by sub-cloning DNA from plasmid pGC511 (see Fig. 1) into the blunt-ended SmaI site of M13mp8. The plasmid pGC511 was cut with BglI and the 1097 base pair fragment containing most of the B gene was extracted from a low melting point agarose gel according to the procedure of Yang et al. (19). This fragment was cloned intact or cut with TaqI, AluI or AluI plus DdeI and the mixtures ligated with dephosphorylated SmaI cut M13mp8. RsaI fragment clones were made from an RsaI digest of pGC511. Prior to ligation with M13mp8, DNA fragments with protruding 5' ends were converted to blunt ends by repair synthesis with the Klenow fragment of DNA polymerase I, while protruding 3' ends were blunted using the 3' exonuclease activity of T4 DNA polymerase (20). Blunt end ligations were carried out according to the procedure in the BRL manual. Hanahan's (21) procedure was used to transfect JM103 with the ligation mixtures.

DNA sequence determination and analysis

Dideoxy sequencing: The DNA sequence from the start of the B gene to the BglI site near the 3' end of the gene was obtained from M13 clones using a modified version of the dideoxynucleotide chain termination method of Sanger (22) as described in the BRL sequencing manual. The modifications of this procedure were taken from Biggin et al. (23) and were as follows: 1) [α -³⁵S]dATP at a specific activity of 400-600 Ci/mM was substituted for [α -³²P]dATP, 2) the buffer for the sequencing reactions consisted of 5 mM Tris-HCl, pH 8.0/2.5 mM MgCl₂, and 3) the chase contained all four dNTPs at 0.5 mM instead of just dATP. Sequencing samples were routinely run on 40 cm long 8% polyacrylamide gels in a buffer containing 133 mM Tris-OH, 44 mM boric acid and 2.5 mM Na₂ EDTA at pH 8.9. To extend the sequences of certain clones up to 550 bp, samples were also analyzed on 6 or 8% gels 60 or 90 cm long.

Maxam and Gilbert sequencing: The plasmid pRA600 (see Fig. 1) was used as a source of DNA restriction fragments for sequencing from the right most BclI site in B to the 3' end of the B gene by the chemical modification method of Maxam and Gilbert (24). All fragments to be sequenced were treated with calf intestinal alkaline phosphatase and end labelled. The 5' ends were labelled using T4 polynucleotide kinase as described by Maxam and Gilbert (24), while the 3' ends of the same fragments were labelled using the Klenow fragment of E. coli DNA polymerase I as outlined in Maniatis et al. (20). Sequencing was accomplished using the Maxam and Gilbert method (24) with a formic acid modification reaction similar to that described in Maniatis et al. (20) in which ^{32}P labelled DNA in H_2O was reacted with 88% formic acid at 20°C . The first 60 bp of each fragment were resolved on 20% polyacrylamide denaturing gels 40 cm long. Bases 35 to 200 were resolved on 40 cm long 8% polyacrylamide buffer gradient gels run at 1200V as described by Biggin et al. (23) using a 1x to 5x buffer gradient. The gel buffer consisted of 100 mM Tris borate pH 8.3 and 2 mM EDTA.

The entire B gene and the A-B intergenic region were sequenced on both strands. In regions which were sequenced by both the dideoxy method and the Maxam and Gilbert method, the sequences were the same. The final sequence was constructed as shown in Figure 1.

RESULTS AND DISCUSSION

Determination of the sequence of the B gene.

Figure 2 shows the complete 939 base pair sequence of the Mu B gene, plus the 38 base pair A-B intergenic region 5' to it. The position of the end of the A gene was assigned to the stop codon in the only open reading frame upstream of the B gene (data not shown). The sequence of the entire B gene and the A-B intergenic region were determined using the sequencing strategy outlined in Figure 1. The open reading frame for the B gene is preceded by the Shine-Dalgarno sequence GAGG at position -9. There does not appear to be a promoter sequence for RNA polymerase prior to the B gene in the A-B intergenic region. This was not unexpected, since the B gene is believed

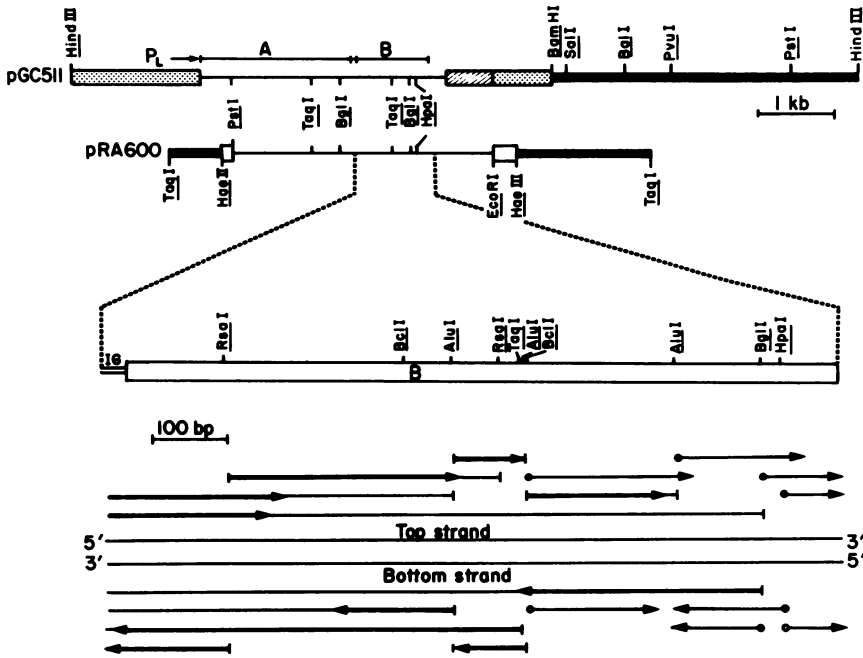


Figure 1. Sequencing strategy for the Mu B gene. The top of the figure shows the plasmids pGC511 and pRA600. The plasmid pGC511 (15) was used as the source of DNA for cloning into M13mp8 for dideoxy sequencing. This plasmid is a derivative of pKC30 (16) carrying a 3.3kb *Bgl*I fragment from Mu₆₂-13/4 (17) that contains the intact Mu A and B genes. The position of the A and B genes transcribed from λ P_L is diagrammed above the restriction map. The plasmid pRA600, used as a source of restriction fragments for Maxam and Gilbert sequencing, carries part of the A gene and all of the B gene on a 3.5kb *Pst*I-*Eco*RI fragment from Mu₆₂ phage DNA cloned into pUC8 (18). Mu sequences are denoted by the thin lines. The heavy lines, stippled boxes, striped boxes and open boxes designate pBR322, λ , IS1 and M13mp8 *lac* sequences respectively. Sequences derived from subfragments ($\overline{\text{---}}$) of pGC511 cloned into M13mp8, and from Maxam and Gilbert sequencing are shown below. The thick lines with arrows indicate the direction and extent of sequencing obtained from each M13 subclone. The thin arrowed lines indicate the extent of sequence obtained by the method of Maxam and Gilbert from restriction fragments labelled at the 5' end (●+) or the 3' end (o+). IG denotes the intergenic region between A and B.

to be transcribed after the A gene from the Mu early promoter. No obvious transcription termination sequences were found in the A-B intergenic region.

The sequence which we present here is identical to the

```

                -40      -30      -20      -10
                -----TAATTAACGTTTAAACAAAATTTAATTACGAGGTTATTTCAG
                ###
1
ATG AAT ATT TCG GAT ATT GGC GCA GGA CTG OGC ACG CTT GTA GAA AAT GAA GAA ACC ACC 60
MET ASN ILE SER ASP ILE ARG ALA GLY LEU ARG THR LEU VAL GLU ASN GLU GLU THR THR
                120
TTT AAA CAA ATT GCT CTT GAG AGC GGC CTT TCT ACC GGA ACT ATC AGT AGT TTT ATC AAT
PHE LYS GLN ILE ALA LEU SER GLY LEU SER THR GLY THR ILE SER SER PHE ILE ASN
                180
GAT AAG TAC AAC GGC GAT AAC GAG OGT GTT TCA CAA ATG CTG CAA OGC TGG CTG GAA AAA
ASP LYS TYR ASN GLY ASP ASN GLU ARG VAL SER GLN MET LEU ARG TRP LEU GLU LYS
                240
TAT CAT GCA GTG GCA GAA CTA CCT GAA COG OCC GCG TTT GTG GAA ACG CAG ACG GTA AAA
TYR HIS ALA VAL ALA GLU LEU PRO GLU PRO PRO ARG PHE VAL GLU THR GLN THR VAL LYS
                300
CAA ATC TGG ACA AGT ATG OGT TTT GCC AGC CTG ACT GAA AGT ATT GCT GTT GTA TGT GGC
GLN ILE TRP THR SER MET ARG PHE ALA SER LEU THR GLU SER ILE ALA VAL VAL CYS GLY
                360
AAT OCT GGT GTG GGC AAA ACC GAA GOG GOC OGT GAA TAT OGC OGC ACC AAT AAC AAT GTC
ASN PRO GLY VAL GLY LYS THR GLU ALA ALA ARG GLU TYR ARG ARG THR ASN ASN ASN VAL
                420
TGG ATG ATC ACC ATT AOG CCA TOC TGT GCC AGC GTT CTG GAA TGT CTT ACT GAA CTG GCG
TRP MET ILE THR ILE THR PRO SER CYS ALA SER VAL LEU GLU CYS LEU THR GLU LEU ALA
                480
TIT GAG CTG GGA ATG AAT GAC GCA CCA OGC OGT AAA GGG COG CTC TOC OGC GOC CTG OGA
PHE GLU LEU GLY MET ASN ASP ALA PRO ARG ARG LYS GLY PRO LEU SER ARG ALA LEU ARG
                540
OGT OGC CTT GAA GGT ACA CAG GGG CTG GTT ATC ATC GAC GAA GCT GAT CAT CTT GGT GCC
ARG ARG LEU GLU GLY THR GLN GLY LEU VAL ILE ILE ASP GLU ALA ASP HIS LEU GLY ALA
                600
GAA GTT CTG GAA GAA CTC OGC CTG TTA CAG GAA TCA ACC OGT ATT GGC CTT GTG CTG ATG
GLU VAL LEU GLU GLU LEU ARG LEU LEU GLN GLU SER THR ARG ILE GLY LEU VAL LEU MET
                660
GGA AAT CAC OGG GTT TAT TCA AAT ATG ACG GGG GGT AAC AGA ACG GTT GAA TTT GOC OGT
GLY ASN HIS ARG VAL TYR SER ASN MET THR GLY GLY ASN ARG THR VAL GLU PHE ALA ARG
                720
CTG TTT TOC OGT ATT GCA AAG OGC ACT GCA ATT AAT AAA ACC AAA AAA GOC GAT GTA AAA
LEU PHE SER ARG ILE ALA LYS ARG THR ALA ILE ASN LYS THR LYS LYS ALA ASP VAL LYS
                780
GCT AIT GOG GAT GCC TGG CAG AIT AAC GGC GAA AAA GAA CTG GAG TTA TTA CAG CAG AIT
ALA ILE ALA ASP ALA TRP GLN ILE ASN GLY GLU LYS GLU LEU GLU LEU LEU GLN GLN ILE
                840
GOG CAG AAA CCA GGT GOG CTT OGC AIT CTG AAT CAT TCA CTT OGC CTT GCA GOC ATG ACG
ALA GLN LYS PRO GLY ALA LEU ARG ILE LEU ASN HIS SER LEU ARG LEU ALA ALA MET THR
                900
GCT CAC GGT AAA GGT GAG OGT GTT AAC GAA GAT TAT CTG OGT CAG GCT TTC OGT GAA TTA
ALA HIS GLY LYS GLY GLU ARG VAL ASN GLU ASP TYR LEU ARG GLN ALA PHE ARG GLU LEU
                939
GAC CTG GAC GTT GAT ATT TCA ACG CTG CTG OGT AAT TAA
ASP LEU ASP VAL ASP ILE SER THR LEU LEU ARG ASN ###

```

Figure 2. Nucleotide sequence of the Mu B gene and the A-B intergenic region. The nucleotide sequence of the coding strand of the DNA is given from 5' to 3'. The nucleotide positions are numbered beginning with the initiation codon for the B gene. The predicted complete amino acid sequence is shown below the DNA sequence. The GAGG Shine-Dalgarno sequence is underlined and the IHF binding sites within B are indicated by boxes. The termination codons of both the A and B genes are indicated by ###. The underlined amino acid sequence (residues 19-40) shows a strong resemblance to the helix-turn-helix structural motif involved in DNA contact and sequence recognition in a number of DNA binding proteins (see text). The boxed residues are very highly conserved.

sequence of the coding region of the B gene independently determined by H. Preiss and D. Kamp (personal communication). It differs, however, by a single nucleotide from the sequence of the last 111 base pairs of the B gene published by Engler et al. (25). According to Engler et al. (25) the nucleotide at position 842 in our sequence (which corresponds to position 14 in their sequence) is a thymine, while our sequencing shows it to be a cytosine. A cytosine in this position results in an alanine codon, while a thymine results in the incorporation of a valine. Alanine and valine are both nonpolar, small amino acids and the substitution of one for the other would probably not have a drastic effect on the B protein. We are confident of our sequence as only a cytosine in this position will form the BglI site shown to be present by restriction enzyme analysis of pGC511 and pRA600. The difference between our sequence and that of Engler et al. (25) may be due to individual isolate differences: cutting of the DNA with BglI would show if the site is present or not. We have also determined that the unknown nucleotide at position 76 in the sequence of Engler et al. (25) (position 904 in our sequence) is a 5-methylcytosine. Our Maxam and Gilbert sequencing gels show the loss of just one band at this position in all four lanes in the middle of the sequencing ladder. This occurs because 5-methylcytosine reacts so slowly with hydrazine (relative to unmodified cytosine) that the DNA is not cleaved at that position (24). The fact that this position is the internal cytosine in the sequence CCTGG, which is methylated by the dcm methylase (26), explains why this particular cytosine is methylated. Sequence analysis of the opposite strand revealed a guanosine at position 904 verifying the presence of the dcm site.

Amino acid composition of the B protein.

The nucleotide sequence of the B gene predicts that the B protein should contain 312 amino acids and have a molecular weight of 35,061. This is in good agreement with an observed molecular weight of 33,000 displayed on SDS-polyacrylamide gels (27). The amino acid sequence met-asn-ile-ser-asp-ile predicted for the amino terminus of gpB from the nucleotide sequence is identical to that found by direct sequence analysis of the amino

TABLE I
 Predicted amino acid composition of Mu B protein

Amino Acid	gp <u>B</u>		Average Protein (ref.28)	
	Residues	Mole %	Mole %	
Ala	A	27	8.7	8.6
Asx	B(D+N)	30	9.6	(9.8)
Cys	C	3	1.0	2.9
Asp	D	12	3.8	5.5
Glu	E	28	9.0	6.0
Phe	F	8	2.6	3.6
Gly	G	20	6.4	8.4
His	H	5	1.6	2.0
Ile	I	19	6.0	4.5
Lys	K	14	4.5	6.6
Leu	L	36	11.5	7.4
Met	M	8	2.6	1.7
Asn	N	18	5.8	4.3
Pro	P	8	2.6	5.2
Gln	Q	12	3.8	3.9
Arg	R	28	9.0	4.9
Ser	S	17	5.4	7.0
Thr	T	22	7.0	6.1
Val	V	18	5.8	6.6
Trp	W	4	1.3	1.3
Tyr	Y	5	1.6	3.4
Glx	Z(E+Q)	40	12.8	(9.9)
small aliphatic	A+G	15.1	16.9	
hydroxyl	S+T	12.5	13.1	
acidic	D+E	12.8	11.6	
acid + acid amide	B+Z	22.4	19.8	
basic	K+R+H	15.1	13.5	
hydrophobic	L+V+I+M	26.0	20.2	
aromatic	F+Y+W	5.4	8.3	
charged	D+E+K+R+H	27.9	25.1	

terminus of the purified protein (15). Table 1 shows the amino acid composition predicted from the DNA sequence. These values are in good agreement with the observed amino acid composition of the protein (15). The B protein has an amino acid composition which is quite similar to that of an "average" E. coli protein calculated on the basis of sequence data for 314 molecular species of E. coli proteins (28, see Table 1).

Resemblance to sequence specific DNA binding proteins

A computer comparison of the nucleotide sequence of the B gene and the predicted amino acid sequence of the protein with sequences in gene and protein banks respectively, did not reveal significant homology with any other genes or proteins. However, a manual comparison of the B protein sequence with the set of amino acid sequences of a number of site specific binding

proteins in the region of their bihelical folds showed strong similarities. This helix-turn-helix region is believed to be involved in DNA contact and sequence specific recognition in the phage λ repressor and cro proteins, and in the E. coli CAP protein (see ref. 29). Furthermore, the bihelical DNA binding fold appears to be present in a number of genome regulatory proteins and may represent a common structural motif in DNA binding (30,31). This DNA binding region is often found near the amino terminus of the genome regulatory proteins and the region of amino acids 19-40 of the Mu B protein shows striking similarities to this set of sequences. These similarities are as follows:

- 1) Residues 25, 29 and 35 of gpB are the alanine, glycine and isoleucine residues corresponding to the consensus sequence ala-N-N-N-gly-N-N-N-N-N-Ile(Val).
- 2) Residues 20, 31, 34, 36 and 40 of gpB are represented at least four times in the corresponding position within the group of related proteins.
- 3) As expected, residues at positions 24, 30, 35 and 38 bear large hydrophobic side chains.
- 4) Positions 31, 32 and 36 are occupied by amino acids carrying polar side chains.

We know that the Mu B protein is a DNA binding protein (15), but we have been unable to detect any sequence specific DNA interactions. Further studies will be required to determine whether residues 19-40 near the amino terminus of gpB are involved in the interaction of the protein with DNA.

Frequency of codon usage in the B gene

Ikemura and Ozeki (32) found that the frequency of use of optimal codons is strictly related to the level of production of individual genes in E. coli, but that this correlation does not hold for transposon or phage genes. The frequency of optimal codon use for 6 transposon and phage genes studied by Ikemura (33) averaged 54%. The frequency of optimal codon use for the B gene was 62%, which is slightly higher than the values calculated for other transposon genes.

Possible binding sites for E. coli integration host factor (IHF)

The E. coli integration host factor, which is required for

site specific integrative recombination of bacteriophage λ , is also required for lytic development of bacteriophage Mu (34). Experiments using *E. coli* strains carrying mutations in the himA and himD loci, which encode the subunits of IHF, suggest that IHF is involved in the transcriptional regulation of the Mu early region (R.K. Yoshida and M.M. Howe, N. Goosen and P. van de Putte, personal communication). Furthermore, it appears that the regulation is much more pronounced for the B gene than for the A gene. Although the Mu B gene is transcribed after the A gene from the Mu early promoter, himA mutants express some A but appear to be phenotypically B minus (35; R. Harshey, unpublished results; R.K. Yoshida and M.M. Howe, personal communication). There is no conspicuous transcription termination sequence in the A-B intergenic region which could account for this discrepancy. Mu nuA mutations which overcome the host himA mutations map to two independent sites to the left of the A gene, but their exact positions are not known (36). There are no known nu mutations within, or just prior to, the B gene. It is therefore still a mystery as to how B expression could be regulated independently of A by himA.

The binding site of the integration host factor has recently been identified as AANNNTTGAT (N. Craig and H.A. Nash, personal communication). In an attempt to shed some light on the regulation of gpB synthesis by IHF we searched for IHF recognition sites within the B sequence and in the A-B intergenic region. We found 3 sites which correspond to the pentanucleotide portion of the sequence at positions 115 and 573 (ATCAA) and 911 (TTGAT), but none of the entire sequence. Since the pentanucleotide portion of the IHF recognition site appears to be fairly common (it is found 8 times in pBR322 and 24 times in M13mp7), the significance of this site within the Mu B gene remains uncertain. It may have no biological significance or it may somehow allow IHF to interact directly with the B gene DNA or RNA, affecting either transcription or translation. It has previously been found that IHF activates the synthesis of the λ cII protein at the post-transcriptional level, possibly by binding to a duplex region of RNA (37). It may be that IHF regulates the synthesis of the Mu B protein in a similar fashion.

ACKNOWLEDGEMENTS

We would like to thank Roberta Yoshida and Martha Howe, Nora Goosen and Pieter van de Putte, Nancy Craig and Howard Nash, and Dietmar Kamp and the late Harro Preiss for communication of unpublished results. We would also like to thank Linda Bonis and Dale Marsh for typing of the manuscript and preparation of the figures, and Greg Gloor for helpful suggestions to improve the manuscript. We gratefully acknowledge the computerized comparison of the B gene and protein sequences with other known gene and protein sequences performed by Michael Zucker and Saran Narang, and Dan Haft, respectively. This work was supported by grants from the National Cancer Institute, the Medical Research Council of Canada and the Leukemia Research Fund (Toronto), as well as by grants from the National Science Foundation (PCM8119756) and the National Institute of Health (GM33247-01). J.L.M. is the recipient of a studentship from the Medical Research Council of Canada.

*To whom reprint requests and correspondence should be addressed

REFERENCES

1. Toussaint, A. and Résibois, A. (1983) In Mobile Genetic Elements (ed. J.A. Shapiro) pp.105-158. Academic Press, New York, N.Y.
2. Chaconas, G., Kennedy, D.L. and Evans, D. (1983) *Virology* 128:48-59.
3. Harshey, R.M. (1984) Cold Spring Harbor Symp. Quant. Biol. 49. In press.
4. Chaconas, G., Harshey, R.M., Sarvetnick, N. and Bukhari, A.I. (1981) *J. Mol. Biol.* 150: 341-359.
5. Toussaint, A. and Faelen, M. (1974) *Molec. Gen. Genet.* 131: 209-214.
6. O'Day, K.J., Schultz, D.W. and Howe, M.M. (1978) In Microbiology 1978 (ed. D. Schlessinger) pp. 48-51. ASM Publications, Washington, D.C.
7. Toussaint, A. and Faelen, M. (1973) *Nature (London) New Biol.* 242: 1-4.
8. Faelen, M., Huisman, O. and Toussaint, A. (1978) *Nature (London)* 271: 580-582.
9. Harshey, R. (1983) *Proc. Natl. Acad. Sci. USA* 80: 2012-2016.
10. Wijffelman, C. and Lotterman, B. (1977) *Molec. Gen. Genet.* 151: 169-174.
11. Waggoner, B., Pato, M., Toussaint, A. and Faelen, M. (1981) *Virology* 113: 279-387.
12. Davis, R.W., Botstein, D. and Roth, J.R. (1980) Advanced Bacterial Genetics pp.227-230. Cold Spring Harbor, New York.

13. Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7: 1513-1523.
14. Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) Experiments with Gene Fusions pp.203-204 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
15. Chaconas, G., Gloor, G. and Miller, J.L. (1984) Manuscript submitted.
16. Shimatake, H. and Rosenberg, M. (1981) Nature 292: 128-132
17. Goosen, T., Giphart-Gassler and van de Putte, P. (1982) Molec. Gen. Genet. 186: 135-139.
18. Vieira, J. and Messing, J. (1982) Gene 19: 259-268.
19. Yang, R. C.-A., Lis, J. and Wu, R. (1979) In Methods in Enzymology (ed. R. Wu) 65: 176-182. Academic Press, New York.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
21. Hanahan, D. (1983) J. Mol. Biol. 166: 557-580.
22. Sanger, F., Wicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467.
23. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80: 3963-3965.
24. Maxam, A. and Gilbert, W. (1980) In Methods in Enzymology (eds. L. Grossman and K. Moldave) 65: 499-560 Academic Press, New York, N.Y.
25. Engler, J.A. and van Bree, M.P. (1981) Gene 14: 155-163.
26. May, M.S. and Hattman, S. (1975) J. Bacteriol. 127:768-770.
27. Giphart-Gassler, M., Reeve, J. and van de Putte, P. (1981) J. Mol. Biol. 145:164-191.
28. Dayhoff, M.O., Hunt, L.T. and Hurst-Calderone, S. (1978) In Atlas of Protein Sequence and Structure (ed. M.O. Dayhoff) Vol. 5, Suppl. 3: 363-369.
29. Matthews, B.W., Ohlendorf, D.H., Anderson, W.F., Fisher, R.G. and Takeda, Y. (1983) T.I.B.S. 8:25-29.
30. Matthews, B.W., Ohlendorf, D.H., Anderson, W.F., Fisher, R.G. and Takeda, Y. (1982) Cold Spring Harbor Symp. Quant. Biol. 47: 427-433.
31. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. (1982) Nature 298: 447-451.
32. Ikemura, T. and Ozeki, H. (1983) Cold Spring Harbor Symp. Quant. Biol. 47: 1087-1097.
33. Ikemura, T. (1981) J. Mol. Biol. 151: 389-409.
34. Miller, H.I. and Friedman, D.I. (1980) Cell 20: 711-719.
35. Chaconas, G., Gloor, G., Miller, J.L., Kennedy, D.L., Giddens, E.B. and Nagianis, C.R. (1984) Cold Spring Harbor Symp. Quant. Biol. 49. In press.
36. Yoshida, R.K., Miller, J.L., Miller, H.I., Friedman, D.I. and Howe, M.M. (1982) Virology 120: 269-272.
37. Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I. and Echols H. (1982) Cell 31: 565-573.