IDENTIFICATION OF THE GENES INVOLVED IN THE REPLICATION OF COLIPHAGE 186

A Thesis
Submitted for the Degree of
Doctor of Philosophy
in the
University of Adelaide

by

ARAPAUT VELAYUDHAN SIVAPRASAD,
B.Sc.(Ag.),M.Sc.(Genetics)

Department of Biochemistry

July, 1984

FOR MY PARENTS AND MATTEMMA

C O N T E N T S

CHAPTER 2. MATERIALS AND METHODS

SUMMARY

The aim of this thesis was to identify the phage functions involved in the replication of the coliphage 186 DNA. For this, a fragment of 186 capable of 186-specific replication was isolated and sequenced to give a sequence of 4859 base pairs.

The sequence revealed a series of six potential genes, of which all but one was overlapping with the neighbouring genes by sharing the tetra-nucleotide ATGA. One potential gene was overlapping with its neighbour for 190 base pairs and, in contrast to others, also had a GTG initiator codon.

Alleles of *geneA*, the gene described as essential for replication, actually fell within the reading frames of two potential genes which were subsequently called *LA* and *RA*. In fact, I was able to show that the product of *RA* alone was directly needed for replication from 186 *ori*, but delivery of ribosomes to the beginning of *RA* was essential for the expression of this gene. One role of *LA* in replication was, therefore, envisaged as the delivery of ribosomes to the beginning of *RA*. This hypothesis was supported by the observation that eliminating *LA* without affecting the delivery of ribosomes to *RA* did not abolish replication initiated from the phage *ori* cloned on plasmids. However, the product of *LA* was required for the efficient replication needed for productive phage infection. This was proved by showing that *LAts* phage was deficient in replication at the non-permissive temperature.

The six-base consensus sequence of phage replication origins, CACTAT, was found in the coding region of *RA* at 92% of the chromosome. No significant homology with Phage lambda, phiX174 and *E. coli* origins was seen, and the sequence near 186 origin showed little potential for extensive secondary structures.

Other features noted on the sequence included a promoter at 95% which was later found to be active *in vitro* by M. Pritchard (personal communication). A potential stem loop structure having a ΔG of -11.9 was seen covering the -10 region of this promotor. The consensus sequence for binding of *lexA* repressor of *E. coli* was also seen spanning the -10 region of this promotor. A model for interference to host metabolism by replicating phage DNA could be proposed based on this "SOS Box".

The sequence recognized by the *dnaA* product for binding to the *oriC*, as well as to the transcript of the *dnaA* gene, was found at 22 bases downstream from the start of transcription from p95.

The consensus sequence for the binding of the *E. coli* Integration Host Factor (IHF) was seen at the beginning of *LA*, covering the initiator codon of this gene, and also in the reading frame of a gene preceding *LA*.

During the course of this project three recombinant plasmids which could be used to expose host functions needed for 186 replication were also constructed.

S T A T E M E N T

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to the best of my knowledge and belief, it contains no material previously published or written by another person, except when due reference is made in the text. I hereby give my consent for making this thesis available for loan and for photocopying.

A.V. SIVAPRASAD

ACKNOWLEDGEMENTS

I thank the following people for their help and support during the course of this work:

Professor W.H. Elliott for giving me permission to work in the Department of Biochemistry.

Dr. J.B. Egan for his expert guidance and constant encouragements. I also want to express my gratitude towards him for his understanding nature and the many personal helps. Thank you, Barry.

All members of Lab. 19, past and present, for their suggestions and critical comments, which were invaluable in the successful completion of this project.

Special thanks are due to Dr. Arnie Puspurs for providing the data on *in vivo* translation patterns directed by various plasmids constructed during this work.

Sylvia Francis, Mary Milnthorpe and Ross Foster for excellent technical assistance. I want to thank Sylvia expecially for her motherly affection, which helped me feel at home.

Lesley Crocker for her excellent preparation of the photographs and Hermiene Langes for her excellent typing of this thesis.

Finally, I want to thank Jalaja, without whom this thesis and my life would make little sense.

ABBREVIATIONS

amp	Ampicillin
tet	Tetracycline
cam	Chloramphenicol
A600	Absorbance at 600 nm
bp	base pairs
kb	kilobase
Kd	kilodalton
mg	milligrams
ml	millilitres
M	molar
mA	milliamps
mM	millimolar
MW	Molecular Weight
ng	nanograms
O.D.	Optical Density at 600 nm
p.f.u.	plaque forming units
RF	Replicative Form
ts	temperature sensitive
ug	micrograms
ul	microlitres
V	volts
TCA	Tri-chloro acetic acid

LIST OF FIGURES AND TABLES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

This thesis aims to determine the phage genes necessary for the replication of the bacteriophage 186. Phage 186 belongs to the group of temperate coliphages (Bertani and Bertani, 1971; Younghusband *et. al.*, 1975) and its genome consists of a double stranded DNA of approximately 30 kb in length (Mol. Wt. of 19.7 x 10^6 daltons; Wang, 1967; Mandel and Berg, 1968; Inman and Bertani, 1969; Chattoraj *et. al.*, 1973; Younghusband *et. al.*, 1975). This phage shares several similarities with bacteriophage P2, and these are found in the morphology of phage virions (Bertani and Bertani, 1971) sizes of DNA (Younghusband *et. al.*, 1975), in the nearly identical cohesive ends (Murray and Murray, 1973), at the level of the arrangement of genes having related functions (Lindahl, 1974; Hocking and Egan, 1982a) and at the level of DNA sequence (Younghusband and Inman, 1974; Skalka and Hanson, 1972; Younghusband *et. al.*, 1975). The DNA sequence homology (40%-50%) shared by these two phages is found in the late regions (0 to 65% of phage chromosome from genetic left end) whereas the early region (65-100%) shows little homology (Younghusband and Inman, 1974). Since the replication genes (Hocking and Egan, 1982a; Finnegan and Egan, 1979; Schnos and Inman, 1971; Lindahl, 1970; Geisselsoder, 1976; Lindqvist, 1971) and the origin of replication (*ori*; Chattoraj and Inman, 1973; Schnos and Inman, 1971; Chattoraj, 1978) are located in the areas not showing homology with each other, it suggests differences in the mechanisms and/or specificities of replication.

1.2 SIMILARITIES BETWEEN THE REPLICATION MECHANISMS IN P2 AND 186

Similarities in replication mechanisms do exist between the phages P2 and 186. Thus, both phages replicate as rolling circle and unidirectionally (Chattoraj and Inman, 1973; Schnos and Inman, 1971; Schnos and Inman, 1982). In P2, the replication is initiated from a single strand nick introduced by the action of the *geneA* product (Chattoraj, 1978; Geisselsoder, 1976). The *geneA* product of 186 is believed to function in a similar way.

The end products of P2 replication are closed circles (Lindqvist, 1971) and these are the precursors for the linear DNA which are packaged by the P2 head genes. The viability of P2-186 hybrids (Bradley, *et. al.*, 1975; Younghusband *et. al.*, 1975; Hocking and Egan, 1982d), which use the 186 replication system but P2 head genes (Younghusband *et. al.*, 1975), therefore indicates that the end products of 186 replication are similar to those of P2. Furthermore, general recombination frequencies in these phages are similar and are some 100-fold lower than that of phage lambda (Hocking and Egan, 1982c). The rate of general recombination frequency may reflect the topology of the replicating molecules and therefore this indirectly suggests that the replication of these two phages produces similar forms of DNA molecules which are different from those of lambda.

Finally, the dependence of these phages on the host function *rep* seem to be identical as both phages are unable to replicate their DNA in a *rep3* strain (Calendar *et. al.*, 1970).

The *rep* protein is a DNA helicase (Scott and Kornberg, 1978; Takahashi *et. al.*, 1979; Tessman *et. al.*, 1982) and is non-essential for the replication of the host itself (Denhardt *et. al.*, 1972). Interestingly, perhaps, phages which require this host function all replicate as rolling circle and unidirectionally. The well characterized phages phiX174, M13 and P2 all initiate their rolling circle replication from a single strand nick introduced by a phage encoded protein (Chattoraj, 1978; Geisselsoder, 1976; Eisenberg, *et. al.*, 1977; Francke and Ray, 1972; Henry and Knippers, 1974; Meyer *et. al.*, 1979) and use the *E. coli rep* protein (Denhardt, *et. al.*, 1972; Calendar *et. al.*, 1970, Kornberg, 1980). The exact role of *rep* in the replication of phage 186 is not known but analogy with P2 and phiX174 suggests that its requirement is probably similar to that in these other phages.

1.3 FEATURES OF REPLICATION UNIQUE TO 186

Despite the similarities in replication mechanisms in P2 and 186, certain aspects of the replication are different. Our interest in the study of the replication of phage 186 stems from the three main differences this phage exhibits in its replication behaviour as compared with P2 and lambda.

1.3.1 Need for *dnaA*

Firstly, the phage 186 is unique among double stranded DNA phages in its requirement for the host DNA initiation function *dnaA* (Hooper and Egan, 1981). The *dnaA* gene, mapped

at 83 min. of the *E. coli* genetic map (Bachmann and Low, 1980; Bachmann, 1983) is involved in the initiation of the DNA replication from *oriC* (Hirota *et. al.*, 1970; Fuller and Kornberg, 1983; Kogoma and von Meyenburg, 1983). Reported molecular weight of the product of *dnaA* has varied between 48 and 54 Kd (Hansen and von Meyenburg, 1979; Yuasa and Sakakibara, 1980; Kimura *et. al.*, 1980; Murakami *et. al.*, 1980) and the molecular weight calculated from the DNA sequence is 52,574 (Hansen *et. al.*, 1982). A string of 11 bases, which has been identified as the consensus sequence for the binding of the *dnaA* protein has been found to occur four times in *oriC* (Oka *et. al.*, 1980; Ohmori *et. al.*, 1984; Fuller and Kornberg, 1983) and once at the beginning of the mRNA of the *dnaA* gene itself (Ohmori *et. al.*, 1984; Hansen *et. al.*, 1982). Binding of the gene product to this site is thought to act positively in the initiation of replication from *oriC*, and negatively in the control of transcription of the *dnaA* gene (Fuller and Kornberg, 1983; Ohmori *et. al.*, 1984).

The exact mechanism by which the *dnaA* mediated initiation occurs is unknown, but it is believed that the gene is involved in the synthesis of *ori* RNA (Zyskind, *et. al.*, 1977). This is substantiated by the finding that certain *dnaA* suppressor mutations (*das*) map in the *rpoB* gene of *E. coli* (RNA polymerase subunit; Austin, 1976; Lindahl *et. al.*, 1977; Bagdasarian *et. al.*, 1977). Isolation of suppressors of the *dnaA* mutation and mapping them in the *rpoB* gene indicates a direct interaction between this gene product and the RNA polymerase (Bagdasarian, *et. al.*, 1977; Atlung, 1981). Another class of suppressors of the *dnaA* mutation comprises

the stable DNA replication (*sdrA*) mutants (Kogoma, 1978). These *sdrA* mutants are capable of initiation in the absence of protein synthesis (Kogoma, 1978) whereas the *dnaA* dependent initiation requires concomittant protein synthesis (von Meyenburg *et. al.*, 1979). Furthermore, unlike the normal initation from *oriC* the stable DNA replication induced by the *sdrA* mutation is *recA* dependent (Kogoma *et. al.*, 1981; Torrey and Kogoma, 1982). It was later found that the initiation pathway in *sdrA* mutants was also independent of *dnaA* function, and *oriC* sequence (Kogoma and von Meyenburg, 1983).

Phages lambda and P2 do not require the product of *dnaA* for initiating their DNA replication (Hooper and Egan, 1981; Bowden *et. al.*, 1975; Skalka, 1977). It has been proposed that the product of either gene O or P of lambda, is capable of substituting for the *dnaA* product (Tsurimoto and Matsubara, 1983). Other phages where the requirement for this initiator protein has been studied and found not required include Mu-1, P1 and phiX174 (Hooper, 1979). One instance other than *E. coli* and 186, where involvement of *dnaA* has been found is the RF DNA replication of M13 (Mitra and Stallions, 1976). In this case, however, the need for *dnaA* is indirect as evidenced by the ability of M13 to replicate in an integratively suppressed *dnaA* strain (Mitra and Stallions, 1976). This makes the replication of 186 rather unique and may suggest a functional similarity to *E. coli* DNA replication. However, the mechanism of replication in these replicons differ, in that, coli replication is bidirectional and theta-type (Masters and Broda, 1971; Bird *et. al.*, 1972; Prescott and Kuempel, 1972; Kaguni *et. al.*, 1982) whereas 186 replicates

unidirectionally by rolling circle (Chattoraj and Inman, 1973). As yet unknown similarities in the mechanisms of initiation in *E. coli* and 186 might exist and these may be responsible for the phage's need for the *dnaA* gene product.

As the rolling circle replication is initiated by a nick in the DNA, the absence involvement of *dnaA* in this mode of replication is conceivable if the function of the gene product is in the synthesis of primer RNA. The *dnaA* independent replication of P2 and phiX174 (Hooper, 1979) therefore may possibly be due to their rolling circle mode of replication. Since 186 replicates as rolling circle but still requires *dnaA* suggests that this gene product has some function other than in the synthesis of the primer RNA. The study of replication of 186 may shed some light into the mechanism of action of *dnaA* in initiation. It is, however, possible that the phage's need for this gene product may be indirect as in the case of M13. In fact, preliminary results in this laboratory with integratively suppressed *dnaA* strains indicated that this might be the case. If this is true, one possibility is that the requirement for concomittant host DNA replication is responsible for the apparent need for the *dnaA* gene product in the replication of 186.

1.3.2 Indirect inhibition of DNA replication caused by ultra violet light

Secondly, the replication of the phage 186 is transiently inhibited in cells which have been irradiated with ultra violet (UV) rays prior to infection. This phenomenon is not observed with either P2 or lambda although phage P1 and Mu-1

show this transient inhibition of replication (Hooper and Egan, 1981). Inhibition of DNA synthesis by UV has always been thought to be due to the presence of thymine dimers on the template which physically block the movement of the replication fork (Doudney and Billen, 1961; Masamune, 1976). Since an undamaged 186 replicon is inhibited in a UV-irradiated cell it implies that some UV-induced *trans*-acting signal is involved in this inhibition. This signal can either be an inhibitor of DNA replication or a depletion of some essential function required for initiation, and may be directed towards replicons which show some similarity with the *E. coli* replicon. Comparison of the *E. coli*, 186, P1 and Mu-1 replicons, all showing the UV-delay, suggests that the initiation function *dnaC* (Carl, 1970; Schubach *et. al.*, 1973; Wechsler, 1975; Kobori and Kornberg, 1982b) is the only known common factor required for the replication of all four replicons (Hooper, 1979). A working hypothesis, then, would be that the depletion of the *dnaC* product, due to the repeated initiations at the thymine dimers, was responsible for the transient inhibition of replication. However, increasing the amount of the intracelluar level of *dnaC* protein by introducing a multicopy plasmid carrying the *dnaC* gene (Kobori and Kornberg, 1982a) did not reduce the delay in 186 replication following UV-irradiation (M. Verma, personal communication). The alternative hypothesis based on the UV-induced inhibitor therefore seems more likely to be responsible for the UV-delay of 186. This inhibitor can be either a host protein/molecule or a phage encoded protein. The latter possibility arises as the replication of 186 has

been shown to be switched off at later stages of infection (Hocking and Egan, 1982a), implying that a phage protein was involved in the inhibition of DNA replication. It is conceivable that UV-irradiation of the cell turns on the phage gene, which codes for this hypothetical protein, immediately after infection so that the phage DNA is inhibited from replication until the block is lifted. This, however, will not explain the delay in replication the host DNA experiences after UV-irradiation. It is hoped that the study of 186 DNA replication may give new information on the UV inhibition of DNA replication, even if it is only about the inhibition of 186 DNA replication.

1.3.3 Multiple initiation

Thirdly, phage 186 is uniquely capable of multiple initiations from its *ori* (Chattoraj and Inman, 1973; Schnos and Inman, 1982). Initiations from the origins of *E. coli*, lambda and P2 are strictly controlled, and repeated initiations do not generally occur (Pritchard, 1978; Schnos and Inman, 1982). Mutations which map in or near *oriC* are known and these mutations result in more frequent initiations from *oriC* (Soll, 1980). No such mutation has so far been isolated in lambda or P2, and in these replicons the successive initiation apparently follow the completion of the previous round of replication (Schnos and Inman, 1982). The factor that controls initiations from these phage origins as well as *oriC* could be either a structural feature of the DNA sequence near *ori* or a phage or host encoded protein(s). Alternatively, limiting amounts of initiator proteins can possibly bring

about single initiations from the *ori*. In lambda this does not seem to be the case, as simultaneous initiations from two *ori* carried on the same DNA have been observed (Schnos *et. al.*, 1982). However, repeated initiations from the same *ori* are seen in the presence of 2mM caffeine (Schnos and Inman, 1982). It is not known whether this reinitiation is caused by the DNA destabilizing effect of caffeine (Ts'O *et. al.*, 1962), or due to a relaxation of the control on the frequency of initiations (Schnos and Inman, 1982). The apparent absence of control on initiation from 186 *ori* is worth studying, as knowledge regarding the control on initiation, and consequently control on replication and cell division, is limited at present. Characterizing the genes involved in replication of 186 may serve to understand the control of initiations of replication in this phage.

1.4 KNOWN FACTS ABOUT THE REPLICATION OF 186

1.4.1 Origin of replication

The only information on the origin of replication and the nature of the replicating molecules came from the work of Chattoraj and Inman (1973). Their observations are summarized below:

Electron micrography (EM) of replicating, partially denatured 186 molecules have located the origin of replication (*ori*) at 92.9 ± 1.8% of the chromosome. The molecules have branched circles and most of them are with a protruding piece of single stranded DNA out of the branch point. It is not

clear what these single stranded 'whiskers' represent. Branches are connected to the circles by single strands, and the branches migrate unidirectionally to the right of the phage DNA. The replication of 186, therefore is unidirectional. The branches originate from a single point on the circle (92.9 ± 1.8% from the genetic left end of the chromosome) and so the replication is similar to lambda and P2 with respect to the number of *ori* sites (Schnos and Inman, 1970; Schnos and Inman, 1971). Most of the molecules are monomeric circles although 3% (of about 450 molecules observed) are dimeric circles. Branch lengths up to 99% of the molecules have been recorded for most, but in five out of the 450 circles observed branch lengths more than that of the length of circles have been found.

Evidence for repeated initiations came from the observation of 2 to 15% of molecules possessing two branch points. In these instances both the branches are found to move to the right, the direction of replication. Although from among the molecules observed under EM these double branched molecules appeared to be exceptions rather than rule, their presence suggests that termination of one round of replication is not a pre-requisite for subsequent initiation. As mentioned earlier in this chapter, this character is unique to 186 compared with the other well known temperate coliphages. Phage T4 replication is the only other known instance where new initiations are observed before the completion of the previous round of replication (Delius, *et. al.*, 1971).

1.4.2 Association of replicating DNA with host components

Replicating DNA usually associate with cell membrane as evidenced by the attachment of the *E. coli* DNA (Hendrickson *et. al.*, 1981; Jacq *et. al.*, 1980; Yamaki *et. al.*, 1980), of lambda (Valenzuela, 1975; Klein *et. al.*, 1980) and of P2 (Ljungquist, 1973; Geisselsoder, 1976) to the *E. coli* cell membrane prior to replication. The attachment of DNA to cell membrane was first predicted by Jacob *et. al.*, (1963) in their replicon hypothesis, postulating that such attachment would be necessary for the segregation of daughter DNA molecules as well as for control on initiation. Two membrane proteins B and B' which bind specifically to two separate sites in or close to *oriC* have been isolated (Jacq *et. al.*, 1980). Indications that the binding site of B' protein regulates initiations from *oriC* come from the finding that the absence of one of these two sites results in unidirectional replication (Jacq *et. al.*, 1980), whereas the presence of both sites results in bidirectional replication (Messer *et. al.*, 1980). Initiations from the *ori* of lambda may also be regulated by the membrane attachment as suggested by the finding that *geneO*, one of the initiator proteins of lambda, is membrane bound (Klein *et. al.*, 1980).

To date there has been no compelling evidence to say that replicating 186 DNA is associated with cell components. However, analogy with P2 may suggest that replication of 186 also follows its attachment to the cell membrane.

1.4.3 Kinetics of replication

The product of *geneA* is required for the replication of 186 DNA (Hocking and Egan, 1982a). Replication of the phage DNA commences about 15 minutes after infection, or induction, and reaches a maximum at about 35 minutes after which the rate drops (Hocking and Egan, 1982a). The drop in the rate of DNA synthesis seems to be due to a late function activated by the *geneB* product of the phage, as evidenced by the absence of this drop during the replication of phage which carried an amber mutation in *geneB* (Hocking and Egan, 1982a). Three conceivable reasons for the switch off of DNA replication are, cell lysis, DNA packaging and an inhibition of replication caused by a phage encoded protein. The third alternative, which if true, will be an interesting aspect of DNA replication to study. This can be distinguished from the other two by the use of a double mutant which neither lyses the cell nor packages the DNA. Experiments using single mutants of lysis or head genes have shown that the inhibition of DNA replication prevails (Hocking and Egan 1982a).

1.4.4 *GeneA*, the only known replication gene in 186

Genetic mapping of amber allels in *geneA* (Hocking and Egan, 1982c; Finnegan and Egan, 1979) places this gene in the region of the chromosome between 83.8 and 87.0%. All the eight amber alleles identified as replication deficient have been found to reside in *geneA* which indicates that this gene is probably the only replication gene in this phage. This is in contrast to the situation found in lambda and P2 where two replication genes each, *genes O* and *P* for lambda (Ogawa and

Tomizawa, 1968; Takahashi, 1975) and *A* and *B* for P2 (Lindahl, 1970; Lindqvist, 1971), have been identified. It has also been suggested (Hocking and Egan, 1982a) that the presence of only one replication gene may be responsible for the unique dependence of this phage for the host DNA initiation functions for *dnaA* and *dnaC* (Hooper and Egan, 1981).

In vivo transcription pattern (Finnegan and Egan, 1981) together with genetic mapping of amber alleles (Finnegan and Egan, 1979) and the determination of the size of the *geneA* protein (Nogare, 1980) helped to propose that this gene spanned the area of DNA between 83.8 and 87.0% of the chromosome. The *ori* is mapped at 92.9 ± 1.8% of the chromosome and therefore more than 1.5 kb of DNA intervenes the replication gene and *ori*, a situation contrasting to what is observed in phages lambda and phiX174, where the respective origins reside in the coding region of the replication genes themselves (Langveld, *et. al.*, 1978; Schnos and Inman, 1970). There is evidence suggesting that the *geneA* of P2 overlaps with the P2 *ori* (Schnos and Inman, 1971) and so in this phage too the spatial arrangement of replication gene and *ori* conforms with the general pattern. In M13 the site of initiation of replication is outside the coding region of the replication gene, *gene2*, but very close to it (approx. 200 bases; Meyer *et. al.*, 1979). Phage T7 provides another instance where the *ori* is found outside the coding region of the replication gene. In this case, the primary origin has been mapped to a 187-bp fragment that separates two of the replication genes, *1* and *1.1* (Fuller *et. al.*, 1983). The phage 186 therefore exhibits a difference in this aspect of DNA replication too.

However, a possibility exists that one or more genes, situated to the right of 87% and overlapping with *ori* is involved in replication, and that such gene(s) escaped detection during the initial isolation and characterization of amber mutants.

1.4.5 *GeneA* probably acts in *cis*

Mutations in *geneA* are characterized by the very poor complementations with mutations in other genes, which probably reflects the *cis*-action of this gene product (Hocking and Egan, 1982b). Similar behaviour of the P2 *geneA* has been reported by Lindahl (1970) and for phiX174 *geneA* by Francki and Ray (1972). The *geneA* of P2 and the *geneA* of phiX174 are involved in the formation of the single strand nick that marks the initiation of the rolling circle replication (Geisselsoder, 1976; Francke and Ray, 1972; Henry and Knippers, 1974; Langveld *et. al.*, 1978) and therefore the similarity of 186 *geneA* with them suggests that this gene is involved in the formation of the single strand nick during the initiation of 186 DNA replication.

1.4.6 Host genes needed for 186 replication

In addition to *rep*, *dnaA* and *dnaC*, 186's need for which has been mentioned earlier, the phage requires the products of *dnaB* (Hooper, 1979) and *gyrB* (unpublished observation). Analogy with lambda, P2 and phiX174 (Furth and Wickner, 1983; Hooper, 1979) suggests that the phage 186 is likely to need the products of other *dna* genes for its own replication.

1.5 AIM AND APPROACH

The initial aim in the characterization of 186 replication is to identify all phage functions involved in replication. The emphasis will be on the molecular characterization of the genes and *ori*, and so this thesis aims to identify these functions at the DNA sequence level. It is hoped that the knowledge gained from this study can be used in the development of an *in vitro* replication system to study the phage replication in detail.

The approach used in this work is to identify the minimal length of the phage chromosome which supports 186-specific replication (186 minichromosome) and, then, characterize it by DNA sequence analysis. Identification of the replication genes can then be done by site-directed mutagenesis of the potential genes carried on the minichromosome. An advantage this method has over the conventional genetic analysis of sponataneous and induced mutations is that interference to replication, caused by the presence of phage morphogenesis and control genes, will be minimal so that the effect of mutational inactivation of genes can be easily interpreted. Futhermore, polar mutations, which can escape detection by using genetic methods alone, may become apparent if the DNA sequence was available. Identification of *ori* will be facilitated by having the DNA sequence by way of exposing the secondary structures and repeats which usually characterize the origins of replication (Hobom, 1981), and by providing a framework for locating the exact *ori* by using *in vitro* or *in vivo* means.

CHAPTER 2

MATERIALS AND METHODS

2.1 BACTERIAL STRAINS

Bacterial strains used in this thesis are listed in Table 2.1.

2.2 BACTERIOPHAGE STRAINS

Phage 186 strains used were 186*cItsp* (temperature sensitive *cI* repressor; Woods and Egan, 1974; Baldwin *et. al.*, 1966) for preparing DNA, and 186 *vir1* (insensitive to 186 immunity; Woods, 1972) for testing 186-sensitivity. The 186*Aam* phages were obtained by heat induction of the appropriate lysogens (see Table 2.1). Phage P2*vir22* (Bertani, 1975) was used for testing sensitivity to P2 infection. All these strains and phage P1kc were obtained from the laboratory collection of Dr. J.B. Egan. M13mp7, M13mp8 and M13mp9 (Messing, *et. al.*, 1981; Messing and Vieira, 1982) were used for cloning the restriction fragments for sequencing.

2.3 PLASMIDS

Relevant features of the various plasmids constructed are given in Table 2.2. Plasmids pBR322 (Bolivar, *et. al.*, 1977) and pBR329 (Covarrubias and Bolivar, 1982) were prepared from cultures of E2106 and E2137, respectively.

Table 2.1 - Bacterial Strains*

Strains No.	Other No.	Genotype	Relevant Characters	Reference/ Source
E251	W3350	*gal,strA*	*Su-*	Finnegan, J.
E252	W3350	E251(186*cItsp*)	*Su-* lysogen of 186*cItsp*	Finnegan, J.
E508	C600	*tonA supE44 thr leu thi*	*Su+*	Hogness, D.S.
E536	W3350	*gal,strA*	*Su-*	Hogness, D.S.
E901	MC1061	*araD139 Δ(ara leu)7697 ΔlacX74 galU galK hsr- hsr+ strA*	*Su-*	
E0941	C2103	*rha-1 his-4 ilv-4 polA1*	*PolA-*	Calendar, R.
E0961	IT1022	*ilvY864::Tn10*	*Tn10* near *rep*	Tessman, I.
E0964	HF4704	*rep3*	*Rep-*	Tessman, I.
E1011	C600	E508(186*cItspAam11*)	*Aam11* lysogen	Hocking, S.M.
E1024	C600	E508(186*cItspAam24*)	*Aam24* lysogen	Hocking, S.M.
E1043	C600	E508(186*cItspAam43*)	*Aam43* lysogen	Hocking, S.M.
E2106	W3350	E536(pBR322)	Source of pBR322	Bolivar, F.
E2137	RR1	RR1(pBR329)	Source of pBR329	Bolivar, F.
E2216	W3350	E536(pEC16)	pEC16 is a clone of 87 to 94% of 186 DNA; *amps*, tet^R	Finnegan, J.
E2249	E536	E536(pEC701)	Minute colonies	This thesis
E2250	E536	E536(pEC702)	Normal colonies	This thesis

E4011	E536	E536 *rep3*	*Rep*-	This thesis
E4012	E251	E251 (186*cItspAts11*)	Lysogenic for *LAts*	This thesis
JM101		*lac pro supE44 thi* F' *traD36 proAB lacIq* ZΔM15	Host for M13 infection	Messing, J.

* Strains E0941, E0961 and E0964 were derivatives of *E. coliC*, whereas all other strains were derivatives of *E. coli K12*

Table 2.2 - Plasmids Constructed during this Work

Plasmid	Construction	Source of DNA	Remarks
pEC701	pBR322 *BamHI*:: 79.6% to 96.0% of 186	186*cItsp* and pBR322	Orientation of insert as p^{tet}-*geneA-186ori*
pEC702	pBR322 *BamHI*:: 79.6% to 96.0% of 186	186*cItsp* and pBR322	Orientation of insert as p^{tet}-*186ori-geneA*
pEC703	As pEC701, but carries *Aam24* mutation	186*cItspAam24* and pBR322	*RA* carries the *Aam24* mutation
pEC704	As pEC701, but carries *Aam43* mutation	186*cItspAam43* and pBR322	*LA* carries the *Aam43* mutation
pEC705	XmnI-96.0% of 186:: pBR322 *EcoRI-AvaI*	pEC701 and pBR322	Carries *RA* but not *LA* This plasmid carries only 186*ori*. Uncharacterized deletions are present on the DNA.
pEC706	186 (94.0%-96.0%):: 186 (XmnI-96.0%):: pBR329 *PstI-BamHI*	mEC5004, pEC701 and pBR329	Carries *RA*, *CP95-LA* fusion gene and *CP93*. Carries only 186*ori*.
pEC707	As pEC701, but *SacI* site of pEC701 deleted		*RA* is mutated at the *SacI* site

pEC708	As pEC701, but *HpaI-NruI* of pEC701 deleted	*RA* is mutated by deleting C-terminal end
pEC709	As pEC701, but *PstI-BamHI* of pEC701 deleted	*CP93*is mutated by deleting its C-terminal end
mEC5004	M13mp93 *PstI-BamHI*:: 94-96.0% of 186	pEC701 and M13mp9

2.4 CHEMICALS

Acrylamide and Bis(N,N'-methylene-bis-acrylamide) were from Sigma Chemical Co.

Agarose and low melting point agarose were from Bethesda Research Laboratories (BRL).

Ammonium persulphate (APS) was from May and Baker and was of analytical grade.

Amine A, Bacto Agar, Bacto peptone, Bacto tryptone, and Yeast extract were from Difco Laboratories.

Ampicillin and Chloramphenicol were purchased from Sigma Chemical Co.

BCIG (5-bromo-4-chloro-3-indolyl-3-DβGalactopyranoside) was from Sigma Chemical Co.

Cesium Chloride was from Bethesda Research Laboratories.

Dithiothreitol (DTT) was from Sigma Chemical Co.

Ethidium Bromide and Acridine orange were from Sigma Chemical Co.

Ethylenediamine tetra-acetic acid (EDTA) was from Sigma Chemical Co.

Ethanol was re-distilled before use and stored at -15°C.

Iso-propyl thiogalactoside (IPTG) was purchased from Sigma Chemical Co.

Mixed Bed Resin was purchased from BIO-RAD Laboratories.

Nucleoside triphosphates were purchased from Sigma Chemical Co.

Phenol was from BDH Laboratories and was re-distilled before use.

Polyethylene glycol (PEG) purchased from the BDH Chemicals Ltd. was used for preparing M13 phages for sequencing. The PEG obtained from Sigma Chemical Co. was used for all other purposes.

Sequencing primer (17*mer*; 5'-GTAAAACGACGGCCAGT-3') was from New England Biolabs.

Sodium dodecyl sulphate (SDS) was from Sigma Chemical Co.

Trizma base and Tris 7-9 were from Sigma Chemical Co.

Tetracycline was a gift from Upjohn Pty Ltd, Australia.

N,N,N',N'-tetramethylethylenediamine (TEMED) was from Eastern Kodak Co.

Other chemicals were routinely obtained from Sigma Chemical Co., BDH Chemicals Ltd, Ajax Chemicals Ltd and May and Baker Ltd and were of either analytical grade or of the highest available purity.

2.5 ENZYMES

Restriction endonucleases were purchased either from New England Biolabs or from Bethesda Research Laboratories. DNA polymerase I (Klenow fragment) was initially purchased from Boehringer Mannheium and later from Biotechnology Research Enterprises of South Australia (BRESA). T4 DNA polymearase and T4 DNA ligase were from Boehringer Mannheium. Calf-intestinal Alkaline Phosphatase (CIP), purchased from Sigma Chemical Co. and column purified according to the procedure of Efstratiadis *et. al.* (1977) was kindly given by Dr. R.H. Symons. RNAse-A was purchased from Sigma and stock solutions

were heated to 80°C to inactivate the DNAses. Bovine Serum Albumin (BSA) was obtained from Sigma Chemical Co. and was acetylated before use according to the procedure described by Gonzalez, *et. al.* (1977). Solutions of BSA at 2mg/ml were prepared in water and stored at -20°C as working stocks.

2.6 RADIONUCLEOTIDES

^{3}H thymidine (25 ci/mMol) and L-[^{35}S] methionine (1200 Ci/mmol were purchased from Radiochemical centre, Amersham, England. $\alpha[^{32}P]$-dATP (2300 Ci/mmol) and $\alpha[^{32}P]$-dCTP (2300 Ci/mmol) were initially obtained as gifts from Dr. R.H. Symons, and later were purchased from BRESA.

2.7 MISCELLANEOUS

Fuji X-ray film was used for autoradiography. Polaroid film from Polaroid film Co. was used for taking photographs of agarose gels. Dialysis tubing was obtained from Union Carbide. Ordinary filter papers and GF/A filters were from Whatmann, and microporous filters from Millipore.

2.8 MEDIA

All media and buffers were prepared with glass double distilled water and sterilized by autoclaving. Stock solutions of amino acids and antibiotics were added from sterile

stock solutions after the media had been autoclaved and cooled to 45°C.

Compositions of media and buffers is given in grams per litre of various components.

2.8.1 L Broth (LB)

Sodium Chloride	-	10 g
Bacto Tryptone	-	10 g
Yeast Extract	-	5 g
Water	-	to 1000 ml

The pH was adjusted to 7.0 before autoclaving.

2.8.2 2 x YT Broth

Yeast Extract	-	10 g
Bacto Tryptone	-	16 g
Sodium Chloride	-	5 g
Water	-	to 1000 ml

The pH was adjusted to 7.0 before autoclaving.

2.8.3 M13 minimal medium

K_2HPO_4	-	10.5 g
KH_2PO_4	-	4.5 g
$(NH_4)_2SO_4$	-	1.0 g
Sodium citrate	-	0.5 g
Water	-	to 1000 ml

After autoclaving and cooling to 45°C added,

20 % $MgSO_4$	-	1.0 ml
20% glucose	-	10.0 ml
1% thiamine HCl	-	0.5 ml

2.8.4 TPGCAA

NaCl	-	0.5 g
KCl	-	8.0 g
NH_4Cl	-	1.1 g
Trizma base	-	12.1 g
KH_2PO_4	-	1.0 g
Sodium pyravate	-	0.8 g
Water	-	900 ml

The pH was adjusted to 7.4 and autoclaved. To 90 ml of this medium were added:

0.16 M Na_2SO_4	-	0.1 ml
1 M $MgCl_2$	-	0.1 ml
0.4 M $CaCl_2$	-	0.25 ml
0.1 mg/ml $FeSO_4$	-	0.1 ml
20% glucose	-	1.0 ml
25% vitamine free Casamino acids	-	4.0 ml

2.8.5 YGC plates

Sodium Chloride	-	10 g
Amine A	-	10 g
Yeast Extract	-	5 g
Bacto Agar	-	15 g
Water	-	to 1000 ml

After autoclaving added,

Glucose (20%)	-	5 ml
$CaCl_2$ (0.4 M)	-	6 ml

YGC plates containing antibiotics were prepared either by adding the appropriate antibiotics from sterile stock solu-

tions to the medium before pouring the plates, or by spreading the antibiotics on plates before use. The final concentrations of the antibiotics were:

Ampicillin	50 ug/ml
Chloramphenicol	30 ug/ml
Tetracycline	20 ug/ml

2.8.6 Z plates

Sodium chloride	-	5.0 g
Amine A	-	10.0 g
Bacto Agar	-	12.0 g
Water	-	to 1000 ml

The pH was adjusted to 7.2 before autoclaving.

2.8.7 M13 minimal plates

The M13 minimal plates were prepared with M13 minimal medium containing 1.5% Bacto agar.

2.8.8 Soft agar

Bacto agar	-	7.0 g
Water	-	to 1000 ml

The pH was adjusted to 7.0, dispensed into 100 ml aliquotes and autoclaved.

2.8.9 YT Soft agar

Yeast Extract	-	5 g
Bacto Tryptone	-	8 g
Sodium Chloride	-	5 g
Bacto agar	-	7 g
Water	-	to 1000 ml

The pH was adjusted to 7.0 before autoclaving.

2.9 BUFFERS

Tris buffers were made from Tris 7-9 and the pH was adjusted with HCl.

TM used for phage preparation contained 10 mM Tris pH 7.1 and 10 mM $MgCl_2$.

10 x TM contained 0.1 M tris pH 8.0 and 0.1 M $MgCl_2$.

10 x TE contained 0.1 M tris pH 8.0 and 1 mM EDTA.

10 x TBE contained 0.89 M tris-borate and 0.01 M EDTA, and was prepared by dissolving at room temperature:

Trizma base	-	108.0 g
Boric acid	-	55.0 g
EDTA	-	9.3 g
Water	-	to 1000 ml

The pH, if varied from 8.3, was adjusted to 8.3 and autoclaved.

10 x TAE contained 0.89 M tris-acetate and 0.01 M EDTA, pH 8.2 and was prepared by dissolving at room temperature:

Trizma base	-	48.2 g
Sodium acetate	-	16.4 g
EDTA	-	3.36 g
Water	-	approx. 600 ml

The pH was adjusted to 8.2 with glacial acetic acid. The buffer was used without autoclaving.

2.10 TECHNIQUES

2.10.1 Culturing of strains

Bacterial cultures were routinely prepared in LB at 37°C. Overnight cultures of JM101 were prepared in M13 minimal medium and subcultured into 2XYT Broth.

When culturing for plasmid preparations or for testing the presence of plasmids, appropriate antibiotics were added to the growth media at the following concentrations:

Ampicillin	50 ug/ml
Chloramphenicol	30 ug/ml
Tetracycline	20 ug/ml

2.10.2 Assaying cultures and phage stocks

Bacterial cultures were assayed by spreading 0.1 ml of appropriate serial dilutions on YGC plates whereas 186 phage lysates were assayed by plating 0.1 ml of appropriate dilutions on Z-plates after mixing with 0.2 ml of indicator bacteria and 3 ml of melted soft agar.

2.10.3 Strain constructions

Phage P1kc was used for generalized transduction described by Miller (1972).

2.10.4 Purification of colonies and plaques

Bacterial colonies were purified by streaking (Miller, 1972). The phage plaques were purified by the overstreaking procedure of Davis *et. al.* (1980).

2.10.5 Replica plating

Replica plating was done as follows by using sterile disks of filter paper (Whatmann no. 2).

The filter paper, cut to fit the petri dish, was carefully placed over the master plate so that no air bubble was trapped in between the paper and agar surface. After it had been fully wetted, the paper was carefully peeled off the and transferred onto the replica plate. The paper was then removed, and both plates incubated overnight. Only one replica was made from each master plate as the colonies seemed to spread and create a smear when used to make second and third replicas.

2.10.6 Cross streaking

Sensitivity of strains to 186 infection was tested by cross streaking loopfuls of bacterial suspensions across phage streaks on YGC plates.

2.10.7 Marker rescue

A lawn of the bacterial strain to be tested was prepared by pouring, on Z-plates, 0.2 ml of the log phase culture (A600 = 0.8 to 1.0) mixed with 3 ml of molten soft agar. After the agar had solidified, the phage lysate (about 10 ul of stock containing 2-4 x 10^8 p.f.u/ml) was placed on the lawn and the plate incubated at 37°C overnight.

2.10.8 Infection and Induction of phage 186

Infections and inductions of phage 186 were done according to the methods given by Hocking and Egan (1982b).

2.10.9 Agarose gel electrophoresis

Agarose gel mix (1% w/v in water) was prepared with 0.089 M tris-acetate and 0.001 M EDTA, pH 8.2 and stored at 65°C. This was used for analytical as well as for preparative gel electrophoresis.

Minigels were prepared by pouring 10 mls of the gel mix on 7.5 cm x 5.0 cm glass microscope slides, with appropriate combs set in place. Samples were loaded with glycerol loading buffer (5% glycerol, 0.04% Bromophenol blue, 25 mM EDTA) and electrophoresed at room temperature at a constant current of 200 mA (ordinary agarose) or 100 mA (low melting point agarose).

Preparative gels were poured either on a polypropylene tray (140 x 110 mm) and run horizontally or poured in a glass sandwich (200 x 200 x 1.5 mm) and run vertically. When the amount of DNA was less than 10 ug, a minigel with a wider well was used for preparative gel electrophoresis.

2.10.10 Staining of gels

Agarose gels were stained with Ethidium bromide (0.0004% w/v in 1 x TAE) and photographed under short wave UV.

When the DNA was to be recovered from gel for cloning or transformation the gel was stained with acridine orange (0.003% w/v in 1 x TAE) for 15 minutes followed by destaining for $\frac{1}{2}$ to 1 hour in fresh TAE.

2.10.11 Estimation of DNA concentrations

The approximate concentrations of DNA solutions were estimated by separating the DNA fragments on minigel and

comparing the intensities of ethidium bromide stained bands with intensities of bands containing known concentrations of *HindIII* digested lambda DNA.

2.10.12 Polyacrylamide gel electrophoresis

(a) Non-denaturing gel

The gel stock (30%) was prepared by dissolving 58 g of recrystallized acrylamide and 1.2 g bis in 200 mls of water at room temperature. This was de-ionized by stirring with 10 g of mixed bed resin (Bio-Rad) for 30 minutes at room temperature. After removing the resin by filtration through scintered glass funnel the gel stock was degassed under vacuum for 30 minutes and stored at 4°C in the dark.

Gels were poured and used on the same day. A 20 x 40 cm gel was prepared by mixing 10 mls of 30% stock solution, 6 mls of 10 x TBE, 385 ul of 25% APS and 96 ul of TEMED and poured into the gel sandwich which had been pre-warmed to 37°C. The gel was allowed to polymerize at 37°C for at least 30 minutes before use. Immediately after removing the comb the wells were rinsed with mono distilled water, and the gel was pre-electrophoresed at 100 V for at least 15 minutes before loading the DNA. Electrophoresis was done at a constant 400 to 500 V. The DNA fragments were always end-labelled before electrophoresis, and the bands were therefore visualized by autoradiography at room temperature.

(b) Denaturing (sequencing) gel

The sequencing gel stock (6% polyacrylamide, 8 M urea in TBE) was prepared by dissolving, at 37°C, 57 g commerical acrylamide, 3 g bis and 480 g urea in 400 mls of double

distilled water. This was de-ionised by stirring 35 g mixed bed resin for ½ to 1 hour and filtered through scintered glass funnel. After adding 100 mls of 10 x TBE, the volume of the gel stock was made upto 1 litre with water and degassed under vacuum for 2 hours. The gel stock was stored at 4°C in the dark and used for a maximum of 2-3 months.

For preparing a 28 x 40 x 0.025 cm gel, 75 ml of the gel stock was mixed with 480 ul of 25% ammonium persulphate (APS; filtered and stored at 4°C) and 120 ul of TEMED, and poured into the plate sandwich which had been pre-warmed to 37°C. After ½ to 1 hour at 37°C the gel was set up vertically on the apparatus. A plastic bag, with the same dimensions as that of the gel and filled with the buffer or water, was kept in place on the front side of the gel with the help of another glass plate. This was done to maintain a uniform temperature across the gel surface, as localized heating near the centre of the gel resulted in 'smiling' of the bands. The gel was pre-electrophoresed for 30 minutes at 1000 V, using TBE as the running buffer. The comb was left in place while pre-electrophoresis, as prolonged electrophoresis without the comb resulted in distortion of the wells. Just before ready to load the reaction mixes the comb was removed and the wells flushed with buffer to remove the unpolymerized acrylamide.

(c) Autoradiography

Gels were autoradiographed at room temperature for 30 minutes to 24 hours, or at -80°C when required to expose for more than 24 hours.

2.10.13 Preparation of high titer stock of Phage 186

High titer stocks of phage 186 were prepared by precipitating the phage particles from 2 liter of the lysate (about 2-3 x 10^{10} p.f.u./ml) by dissolving 100 grams of polyethylene glycol (PEG) and 19 grams of sodium chloride, per litre, and storing overnight at 4°C. The phage pellet was then collected by centrifugation at 8000 r.p.m. for 20 minutes at 4°C and resuspended in 1 to 2 ml of TM (10 mM tris pH 8.0, 10 mM $MgCl_2$). Cesium chloride solution having specific gravities 1.35 and 1.6, respectively, were prepared with TM and a block gradient was formed in a 10 ml Oak Ridge tube by layering 3 ml of the 1.35 solution over 2 ml of the 1.6 solution. The phage suspension was then carefully layered on top of the gradient and centrifuged at 45000 r.p.m. for 90 minutes at 8°C. The phage band in between the two gradient blocks was carefully withdrawn and mixed with an equal volume of saturated CsCl solution made in TM. Two blocks of CsCl solutions (specific gravities 1.6 and 1.35, respectively) were formed above this mix in a 10 ml Oak Ridge tube and centrifuged again at 45000 r.p.m. for 90 minutes at 8°C. The phage band was collected and dialysed against TM, with 3 to 4 changes of the buffer at intervals of 4 to 12 hours. The high titer stock so prepared (approximately 3 x 10^{13} p.f.u./ml) was stored at 4°C.

DNA from the high titer stock was obtained by phenol extraction followed by ethanol precipitation.

2.10.14 Plasmid DNA preparation

Large scale plasmid DNA preparations were done according to the method of Birnboim and Doly (1979) as described by

Maniatis *et. al.* (1982), with the exception that the nucleic acids were precipitated with ethanol. The pellet was then resuspended in 8 ml of TE, and 8 grams of cesium chloride was dissolved in it. From an ethidium bromide solution (10 mg/ml in TE) 200 ul was mixed with the DNA/CsCl solution and centrifuged to equilibrium in a Ti50 rotor at 45000 r.p.m. for 42 hours at 15°C. The DNA bands were visualised in ordinary light and the lower band was collected by draining out through a hole made at the base of the tube. The ethidium bromide was removed by extracting three times with iso-propanol which had been equilibrated with 5 M NaCl, 10mM tris pH 8.0 and 1 mM EDTA. The DNA solution was then dialysed at 4°C in 1 litre of TE. The buffer was changed 3 to 4 times at intervals of 4 to 12 hours.

Small scale plasmid DNA preparations (from 1 to 10 mls of cultures) were basically the same as given by Grosveld et. al. (1981) for cosmid DNA preparations. The DNA plus RNA pellet obtained by this method was dissolved in 20 ul of TE containing 1 ul of RNAseA (10 mg/ml) and incubated at 37°C for 10 minutes. This DNA was used directly for transformations but was gel-purified (section 2.10.17) for restriction digestions.

2.10.15 Phenol Extraction

Redistilled phenol was equilibrated with buffer by mixing 50mls of phenol with 50 mls of 1 M tris pH 8.0 and 5 mg of 5-hydroxy quinoline, and heating in a 65°C oven until the aqueous and organic phases became one (usually took 1 to 2 hours), followed by cooling to room temperature in the dark to

separate the two phases again. This was stored at -20°C as master stock from which aliquotes of 10 mls of phenol phase was taken, mixed with 10 ml of TE (10 mM tris pH 8.0, 0.1 mM EDTA) and let stand at room temperature until the phases separated. The TE phase was then removed, and the whole procedure repeated twice. The phenol, equilibrated and washed in this way, was stored under TE at 4°C as working stock. Fresh working stock was prepared at every 3 to 4 weeks.

Extraction with phenol was done at room temperature by mixing DNA solutions, or phage stocks, with ½ volume of equilibrated phenol and centrifuging at room temperature (for Eppendorf tubes) or at 4°C (when Oak ridge tubes were used). The aqueous phase was withdrawn, and the procedure was repeated, if necessary, until no protein band was visible at the interphase. Finally, the aqueous phase was ethanol precipitated and the DNA resuspended in TE.

2.10.16 Ethanol precipitation

Plasmids and linear double stranded DNAs were precipitated by adding NaCl to 200 mM, 5 ul of tRNA (10 mg/ml) and 2.5 volume ethanol,followed by freezing either at -80°C for 30 minutes or in a dry ice/ethanol bath for 15 minutes. The precipitated DNA was recovered by centrifugation for 15 minutes at room temperature (for small volumes taken in Eppendorf tubes) or at 10K for 15 minutes at 5°C (for large volumes precipitated in 50 ml oak ridge tubes). The inside of the tube was then washed once with 95% ethanol without disturbing the pellet and dried in a vacuum dessicator for 10 minutes. The DNA pellet was dissolved in TE and stored at 4°C.

Single stranded M13 DNA was precipitated essentially like double stranded DNA with the exception that 3M sodium acetate pH 4.8, instead of NaCl, was added to 300 mM.

2.10.17 Gel-purification of DNA

The DNA was run on a mini-agarose gel and after staining with acridine orange and visualization under white light the required band was cut out and electro-eluted.

2.10.18 Electro-elution

The agarose gel slice containing the required DNA fragment was placed inside a dialysis bag (18/32) which had been rinsed with TE. About 400 ul of TE was added, ends tied such that the bag was in a fully inflated state and placed across the path of electric current in a horizontal gel apparatus and electrophoresed at 100 mA for 5 minutes (10 to 20 minutes when the length of DNA fragment was more than 4 kb). The bag was inverted a few times to dislodge any DNA that might be sticking on to the walls, the buffer was taken out by piercing the side of the bag and ethanol precipitated after adding carrier tRNA (5 ul of 10 mg/ml stock; the tRNA stock was prepared by dissolving 20 mg of the commercial powder in 1 ml of water and extracting three times with phenol, followed by one ethanol precipitation. The pellet was dissolved in 1 ml of water and stored at -20°C).

2.10.19 Elution of DNA from low melting point agarose gels

Gel slices containing the required fragments of DNA were cut out of low melting point agarose gel into Eppendorf tubes and approximately 2 volumes of 50 mM tris pH 8.0 + 0.5 mM EDTA

was added. The gel slice was melted by keeping the tube at 65°C for 30 to 40 minutes. After cooling to 37°C the DNA was recovered from the solution by two phenol extractions followed by ethanol precipitation of the aqueous phase.

2.10.20 Elution of DNA from polyacrylamide gel

Gel slices containing the DNA fragments to be eluted were cut out by superimposing the autoradiograph from which positions of the bands had been removed. The DNA was eluted by soaking the gel slices overnight in 400 ul of gel elution buffer (500 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA and 0.1% SDS pH7.6) and the supernatent was ethanol precipitated. No carrier tRNA was used when precipitating the DNA fragments eluted out of polyacrylamide gel slices.

2.10.21 Dialysis

Dialysis bags were prepared by boiling the tubing (18/32) in TE (for DNA) or in TM (for phage 186) for 5 minutes followed by washing the inside of the bags with the same buffer at room temperature. Dialysis tubes were prepared and used on the same day.

Dialysis was done at 4°C in 1 litre of TE (or TM for phage 186) with 3 to 4 changes of the buffer at intervals of 4 to 12 hours. The DNA solution was the ethanol precipitated and resuspended in TE.

2.10.22 Restriction Digestion of DNA

DNA digestions were done at 37°C for 1 hour with the folowing general recipe:

DNA + H_2O	-	15 ul (up to 10 ug DNA)
10 x digestion buffer*	-	2 ul
BSA	-	1 ul
Restriction Enzyme	-	2 ul (0.6 to 20 units)

Reactions were terminated by heating to 65°C for 10 minutes.

* Separate buffers for each enzyme was made according to the assay conditions described in the 1983/84 New England Biolab's catalog, and stored at 15°C.

When needed to digest with more than one enzyme, those which required similar concentration of NaCl were used simultaneously. When the optimal salt concentrations of enzymes differed, the enzyme which required the lowest NaCl concentration was used first and, then, added the other(s) after supplementing the mix with the required amount of NaCl. The DNA was ethanol precipitated and resuspended in the second digestion buffer when the buffers for enzymes differed considerably.

When large amounts of DNA (> 10 ug) were to be digested the reaction volumes, except that of enzyme, were scaled up and digestion continued overnight in a hot room at 37°C.

2.10.23 End-labelling and End-filling

End-labelling to identify small DNA fragments (less than 2 kb) on polyacrylamide gels was done as follows:

^{32}P labelled dATP (or dCTP)	-	4 ul (vacuum dried)
DNA + H_2O	-	8 ul
10 x TM*	-	1 ul
Klenow enzyme	-	1 ul (1 unit)

* TM was not added when the DNA samples were in restriction digestion buffers.

After incubating the above mix at 37°C for 15 minutes added 2 ul of dNTP solution (0.25 mM each of dATP, dCTP, dGTP and dTTP in 5 mM tris pH 8.0 + 1.0 mM EDTA) and incubation continued for a further 15 minutes. The reaction was terminated by heating the tube to 70°C for 10 minutes.

End-filling was essentially similar to end-labelling except that the labelled nucleotide was replaced with 2 ul of the dNTP solution.

2.10.24 Ligation

A typical ligation reaction contained 20 to 100 ng of DNA, 1 mM rATP, 10 mM $MgCl_2$, 10 mM dithiothreitol, 50 mM Tris-HCl pH 7.5 and 0.02 unit (for sticky end ligation) to 0.2 unit (for blunt end ligations) of T4 DNA ligase in 20 ul reaction volume. Ligations were done at 14°C for 1 to 16 hours by keeping the tubes submerged in a water bath. Inactivation of ligase was not done.

For M13 cloning, a vector to donar molar ratio of 1:3 was used. For this, the amount of vector DNA was kept constant at 20 ng per ligation reaction, and the donor DNA was added at

approximately 10 ng/kb of the fragment's length. Ligation reaction was done exactly as described as above.

2.10.25 Preparation of the competent cells

Overnight cultures made in LB were diluted 1:100 into fresh broth and incubated with aeration at 37°C until the A_{600} reached 0.4 to 0.8. The culture was cooled on ice for 20 minutes and the cells sedimented at 5000 r.p.m. for 5 minutes at 4°C. The pelleted cells were resuspended in 1 to 2 ml of ice cold 0.1 M $CaCl_2$ and left on ice for at least 1 hour. Competent cells were prepared and used on the same day.

2.10.26 Transformation

Transformation was by mixing 100 to 500 ng of DNA and 0.2 ml of competent cells in chilled sterile screw capped polycarbonate tubes and keeping on ice for 10 minutes followed by 5 minutes at 37°C and a further 10 minutes on ice. The transformation mix was then diluted by adding 2 ml of LB and incubated for 1 hour at 37°C with aeration. Fractions of 0.1 ml were spread on selective plates and incubated at 37°C.

2.10.27 M13 transfection

M13 transfection was by mixing 1 ul of ligated mix with 0.2 ml of competent cells in chilled sterile glass tubes and keeping on ice for 40 minutes. After heat shock treatment for 2 minutes at 45°C the cells were mixed with 3 ml of molten YT soft agar containing 20 ul of IPTG (24 mg/ml stock solution made in water), 20 ul of BCIG (20 mg/ml stock made in dimethyl formamide) and 0.1 ml of log phase JM101 and poured onto M13 cloning plates. Incubation was at 37°C overnight.

2.11 SEQUENCING OF THE BGLBAM FRAGMENT

2.11.1 Sub-cloning of restriction fragments

The DNA fragments between *SacI* and *BamHI*, and *EcoRI* and *EcoRI* of pEC701 were isolated from agarose gels after staining with acridine orange. The former fragment was digested separately with *HaeIII*, *HpaII* and *TaqI*, the individual fragments isolated from polyacrylamide gels and cloned into M13mp7 RF DNA which had been cleaved with either *AccI* or *HincII*. The *EcoRI-EcoRI* fragment was digested separately with *AluI*, *FnuDII*, *HhaI* and *HpaII*, end-filled and shot-gun cloned into M13mp9 RF DNA which had been linearized with *SmaI*. All other fragments needed were isolated from polyacrylamide gels and cloned into the appropriate restriction sites of M13mp9.

2.11.2 Single strand phage preparation

Overnight cultures of JM101 were diluted into 2 x YT broth and 2 mls each dispensed into sterile screw capped polycarbonate tubes. After 15 minutes of incubation at 37°C with aeration the cultures were infected with M13 phage by tooth picking fresh plaques and incubated at 37°C with aeration for 5 to 6 hours. The cells were sedimented at 6000 r.p.m. for 20 minutes and the supernatent recentrifuged for 10 minutes in an Eppendorf (model no. 5413). The phage particles were then precipitated by adding 270 ul of PEG solution (20% PEG (BDH) and 2.5 M NaCl) into 1 ml of the supernatent and leaving at room temperature for 15 minutes. The phage pellet was collected by centrifugation at room temperature for 5 minutes and resuspended in 200 ul of TE. Phage stocks prepared in this way were stored at 4°C.

2.11.3 Single strand DNA preparation

The phage suspension (200 ul) was phenol extracted with 100 ul of TE-saturated phenol at room temperature. From the aqueous phase 150 ul was carefully withdrawn and ethanol precipitated overnight at -20°C after adding 6 ul of 3M sodium acetate pH 4.8 and 400 ul of ethanol.

2.11.4 Annealing

The universal primer (17mer) was annealed to the template by mixing 1 ul of primer (2.5 ng), 1 ul of 10 x TM and 8 ul of the single strand DNA template. Annealing was done at 60°C for 1 hour and the tubes were slowly cooled to room temperature.

2.11.5 Polymerization Reaction

The sequencing reagents ddNTPs and dNTPs were prepared separately and stored at -20°C. Compositions for these mixes are given in Table 2.3.

Prior to sequencing, the reaction mixes were prepared by combining equal volumes of ddNTPs and dNTPs and dispensing 2 ul each into Eppendorf tubes. The annealed DNA-Primer was mixed with 2 ul of dried down ^{32}P dCTP and 2 ul of label supplement (16 uM dCTP in 5 mM tris pH 8.0 and 0.1 mM EDTA) and 2 ul each dispensed into four reaction tubes. The Klenow enzyme solution (1 unit/ul) was diluted just before use by mixing 6 ul with 95 ul of 1 x TM, and 2 ul each was dispensed onto the side walls of the reaction tubes. The sequencing reaction was commenced by a quick centrifugation to bring down the enzyme solution into the reaction mix. After exactly 15

Table 2.3 - Sequencing Reagents

dNTP mix

	A'	C'	G'	T'
dATP	11 uM	153 uM	215 uM	215 uM
dGTP	251 uM	153 uM	16 uM	215 uM
dTTP	215 uM	153 uM	215 uM	16 uM

The sequencing reagents were prepared by mixing the required amounts of dNTP solutions from 20 mM stocks (in water), and stored at -20°C.

Symbols A', C', G' and T' refer to the respective mixes containing the rate limiting deoxy-ribo nucleotide triphosphate.

All dNTP mixes were made in 5 mM Tris pH 8.0 and 0.1 mM EDTA.

The ddNTP solutions were made in water.

B. ddNTP solutions

ddATP - 0.5 mM

ddCTP - 0.1 mM

ddCTP - 0.3 mM

ddTTP - 0.8 mM

minutes at 37°C, 2 ul of dNTP solution (0.25 mM each of dATP, dCTP, dGTP and dTTP in 1 x TE) containing 1/32 dilution of the Klenow enzyme solution was added to each tube. The reaction was continued for a further 15 minutes and then terminated by adding 4 ul of freshly prepared formamide loading buffer (95% deionized formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10 mM EDTA and 10 mM NaOH). Just before loading onto the gel the reaction mixes were heated to 100°C for 2.5 minutes, and 0.5 ul each loaded onto gel. The gels were run at a constant 1500V (gel temperature varied from 40 to 50°C) until the bromophenol blue dye migrated to about 2 cm from the bottom of the gel.

2.11.6 Fixing the gel and autoradiography

After removing the glass plate the gel was immediately washed with 10% acetic acid. The washing was continued for about 15 minutes until all urea had been removed from the gel. The gel was then baked in a 110°C oven for 45 minutes and autoradiographed overnight at room temperature.

2.11.7 Sequencing with specific primers

Specific primers were prepared by isolating the required DNA fragments either from the RF DNA of the appropriate M13 clones or from pEC701 DNA. Annealing was essentially same as for normal sequencing except that the DNA-primer mix was heated to 100°C for 3 minutes before incubating at 60°C. Sequencing was done exactly in the same way as described above.

2.11.8 Reversing the clones

Orientations of insert in M13mp9 clones were reversed by isolating the insert after digesting the RF DNA with *HindIII* and *BamHI*, and cloning this into M13mp8 which has been linearized with these two enzymes.

Chapter 3

CONSTRUCTION AND CHARACTERIZATION OF THE 186 MINICHROMOSOME

3.1 INTRODUCTION

The initial aim in the determination of replication genes was to identify a 186 DNA fragment that would support phage-specific replication. Such a minichromosome should carry all the phage functions needed for replication and so the identification of these functions at the DNA sequence level would be facilitated by its use.

The two phage functions known to be required for replication of 186 are *geneA* (Hocking and Egan, 1982a) and the origin of replication (*ori*; Chattoraj and Inman, 1973). Genetic mapping of amber alleles of *geneA* indicated that this gene spans the 83.8% to 87.0% of 186 chromosome (Finnegan and Egan, 1979; Hocking and Egan, 1982c). The *ori* has been physically mapped to lie at 92.9 ± 1.8% of the genome (Chattoraj and Inman, 1973) and so the DNA between 79.6% and 96% of the chromosome, defined by the *BglII* and *BamHI* sites (Saint and Egan, 1979; Finnegan and Egan, 1979; Hocking and Egan, 1982c), must contain both these functions. By constructing a minichromosome the question could be asked whether this piece of DNA, termed BGLBAM fragment, was capable of 186-specific replication. Involvement of other phage functions, if any, in replication might become evident once the minichromosome had been characterized.

This chapter describes the construction and characterization of a plasmid (pEC701) which is capable of 186-specific replication.

3.2 METHODS

This section describes the procedures employed in the construction and characterization of plasmids, pEC701 and pEC702. See Chapter 2 for detailed descriptions of the techniques involved.

3.2.1 Construction of pEC701 and pEC702

The plasmid pBR322 DNA was prepared from 500 ml of overnight culture of E2106 (=E536(pBR322)) by the large scale plasmid preparation method and the DNA was purified by CsCl-EtBr centrifugation. From this, 10 ug of DNA was cut with *BamHI* to prepare the vector DNA.

Phage 186 DNA was prepared from a high titre stock of 186*cItsp* phage and digested sequentially with *BglII* and *BamHI* and the 4.9 kb band isolated from low melting point agarose gel. About 100 ng of this fragment was ligated with 100 ng of vector DNA and transformed into E536. Colonies were selected for *AmpR* and screened for *TetS* by spotting on YGC + Amp + Tet plates. Miniplasmid preparations from ten colonies were run on agarose gel to determine the relative sizes of uncut plasmid DNA, and to gel purify the DNA for restriction analysis. The DNA was digested with *EcoRI* to determine the relative orientations of inserts.

3.2.2 Construction of pEC703

The 186*cItspAam24* phage lysate was prepared by heat induction of 100 mls of E1024 (= E508 (186*cItspAam24*)) followed by precipitation of the phage particles with poly-ethylene glycol, as described for the preparation of the high

titer stock. The DNA was prepared by phenol extraction of the phage pellet and, after gel purification to separate away from contaminating *E.coli* DNA, was double digested with *BglII* and *BamHI*. The 4.9 kb fragment, isolated from agarose gel, was ligated with *BamHI* digested pBR322 DNA and transformed into an Su^+ strain (E508). One minute colony was purified and the plasmid analysed by restricting with *EcoRI*.

3.2.3 Construction of pEC704

The construction of pEC704 was similar to that of pEC703 except that the source of phage was the strain E1043 (= E508 (186*cItspAam43*)).

3.2.4 Inactivation of p^{tet}

The plasmid pEC701 was prepared from 500 ml culture of E2249 (= E536 (pEC701)) and purified by CsCl-EtBr centrifugation. The DNA (approximately 5 ng) was then digested with *ClaI* and end-filled. After separating on 1% agarose minigel the linear form was isolated, ligated and used to transform E536.

3.2.5 Construction of E536 *rep3*

Phage P1kc raised on E0961 (*ilv- zid::Tn10 rep+*) was used to transduce *Tn10* into E536. One colony which was tet^R *ilv-* was purified by streaking. The *ilv+* gene was transduced into this by using P1kc lysate prepared on E0964 (*ilv+ rep3*). Transductants were screened for 186 and P2 resistance by cross streaking against lysates of 186*vir1* and P2*vir22*.

3.3 RESULTS

3.3.1 CLONING OF THE BGLBAM FRAGMENT

The restriction map of the 186 chromosome is given in Fig. 3.1. The pattern of fragments produced by double digestion of 186 DNA with *BglII* and *BamHI* is given in Fig. 3.2. The BGLBAM fragment, represented by the 4.9 kb band (shown by arrow) was isolated from 186*cItsp* phage DNA and cloned into the *BamHI* site of pBR322. The fragment was therefore inserted behind the promoter for the tetracycline resistance gene (p^{tet}) of the plasmid (Sutcliffe, 1979; Siebenlist *et al*, 1980).

3.3.2 STRUCTURAL CHARACTERIZATION OF CLONES CARRYING THE BGLBAM FRAGMENT

The restriction enzymes *EcoRI* and *SacI* have one asmmetrically placed site each on the BGLBAM fragment (Fig. 3.1). Presence of the insert as well as its orientation relative to p^{tet} could therefore be determined by the characteristic pattern of fragments generated by these enzymes in combination with one which had a known site on pBR322. The patterns of fragments generated on digesting the plasmid DNAs with *EcoRI* are given in Fig. 3.3. The plasmid which had the orientation p^{tet}-*geneA*-186*ori* was named pEC701 and the one that had p^{tet}-186*ori*-*geneA* was named pEC702 (Fig. 3.4). Identity of the insert in pEC702 was also confirmed genetically by the ability of this plasmid to marker rescue the infecting 186*cItspAam11* phage, whereas genetic testing was not possible with pEC701 because of the 186 resistant (186^R) character of the colonies. The 186^R nature of strains

Fig. 3.1 Genetic Map of 186 Chromosome

Locations of the restriction sites *BamHI* and *PstI* are as given by Finnegan and Egan (1979).

Position of *BglII* given by Saint and Egan (1979) had subsequently been changed from 81% to 79.6%.

The position of the *SacI* site has been obtained from unpublished results in this laboratory.

Map positions of genes and their functions are given by Hocking and Egan (1982b, c and d), whereas the order of alleles of *geneA* is given by Finnegan and Egan (1979).

The *ori* has been located by Chattoraj and Inman (1973) to lie at 92.9 ± 1.8%. This range is shown in the detailed map of the *BglII* to *BamHI* fragment.

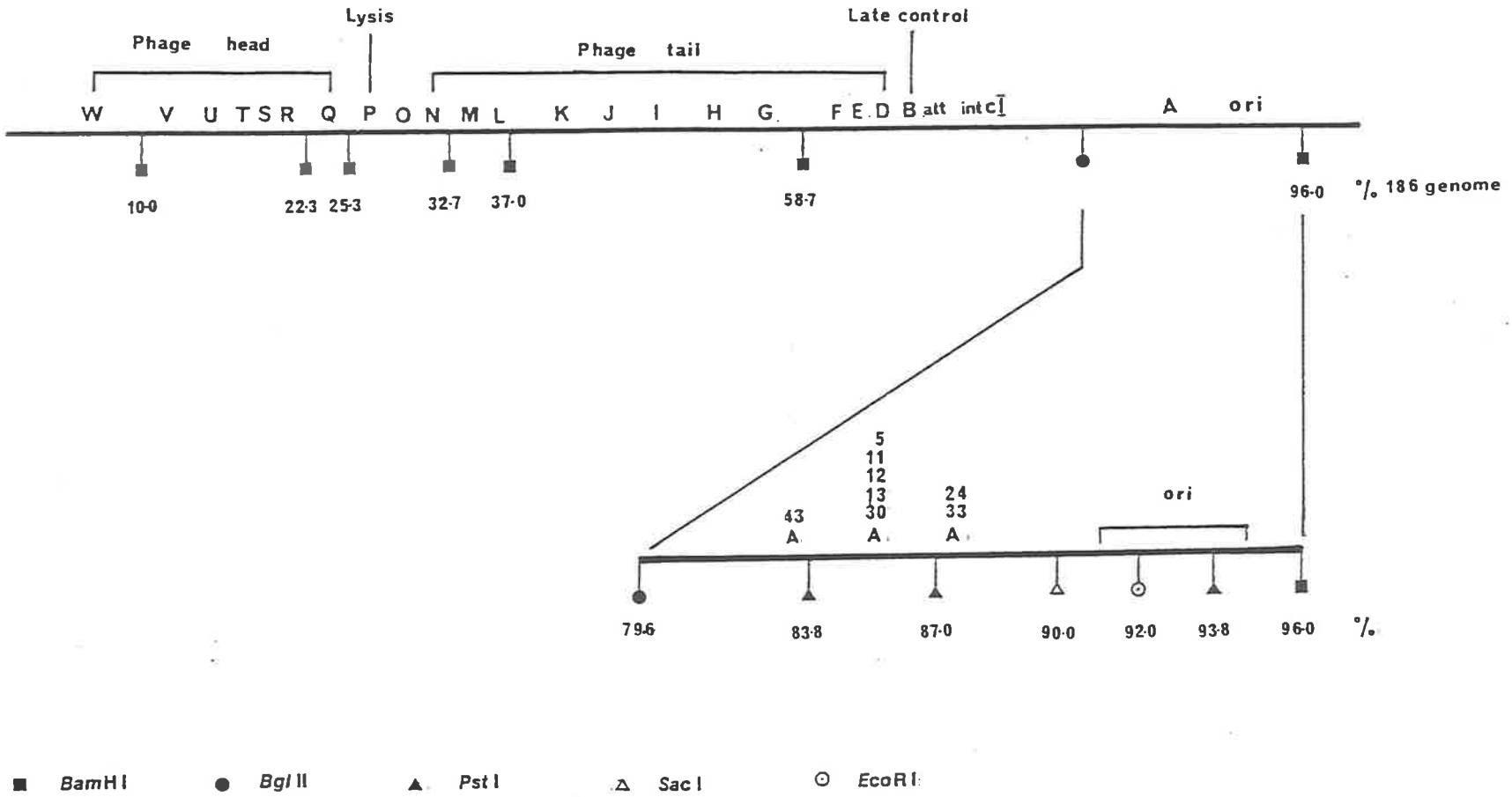

Lysis
Late control
Phage head
Phage tail
W V U T S R Q P O N M L K J I H G F E D B att int cI A ori
10·0
22·3
25·3
32·7
37·0
58·7
96·0
% 186 genome
5
11
12
13
30
A
43
A
24
33
A
ori
79·6
83·8
87·0
90·0
92·0
93·8
96·0
%
BamH I
Bgl II
Pst I
Sac I
EcoR I

Fig. 3.2 Fragments generated by double digesting 186 DNA with *BamHI* and *BglII*

Tracks:

1. 186*cItsp* DNA digested with *BamHI* and *BglII*
   Fragments' sizes (kb): 6.5, 6.3, 4.9, 4.2, 3.7, 2.2.
   The band shown by arrow represents the *BglII* ti *BamHI* DNA fragment (BGLBAM fragment).

2. Lambda DNA digested with *HindIII*. Sizes in kilobases are marked alongside the track.

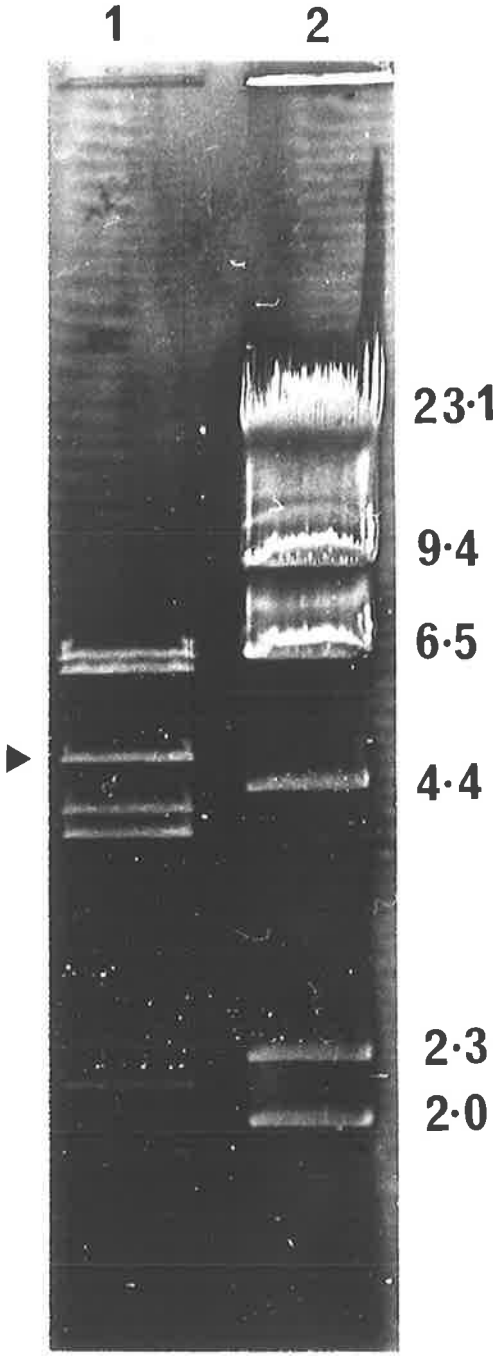

1
2
23·1
9·4
6·5
4·4
2·3
2·0

Fig. 3.3 Fragments generated by digesting pEC701 and pEC702 with *EcoRI*

Gel-purified plasmid DNAs were digested with *EcoRI* and fragments separated on 1% agarose gel. Sizes (kb) are marked alongside. The sizes of fragments agree with the predicted sizes if the orientation of inserts in these plasmids are as shown in Fig. 3.4.

Tracks:

1. Lambda DNA, digested with *HindIII*
2. pEC702 DNA, digested with *EcoRI*
3. pEC701 DNA, digested with *EcoRI*

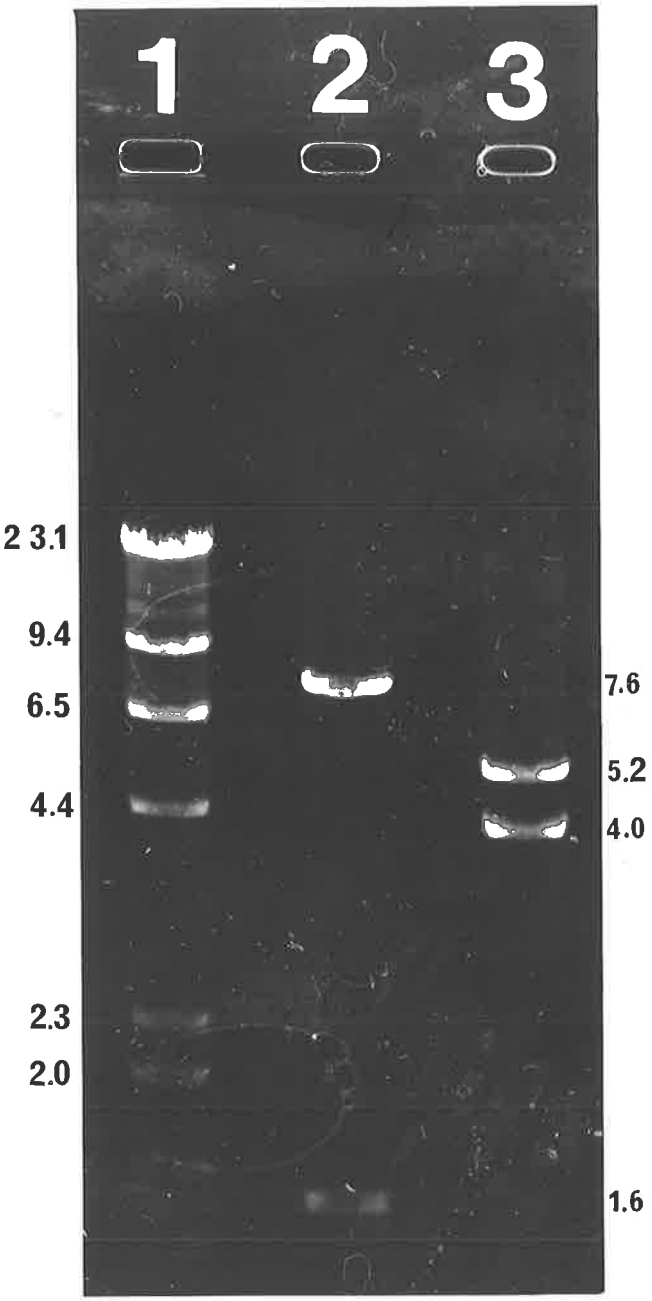
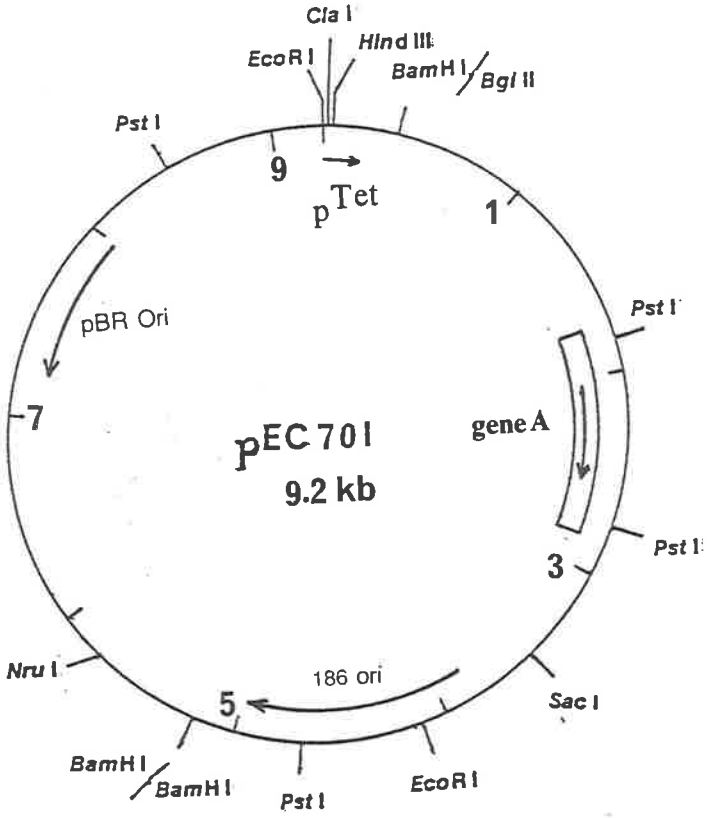
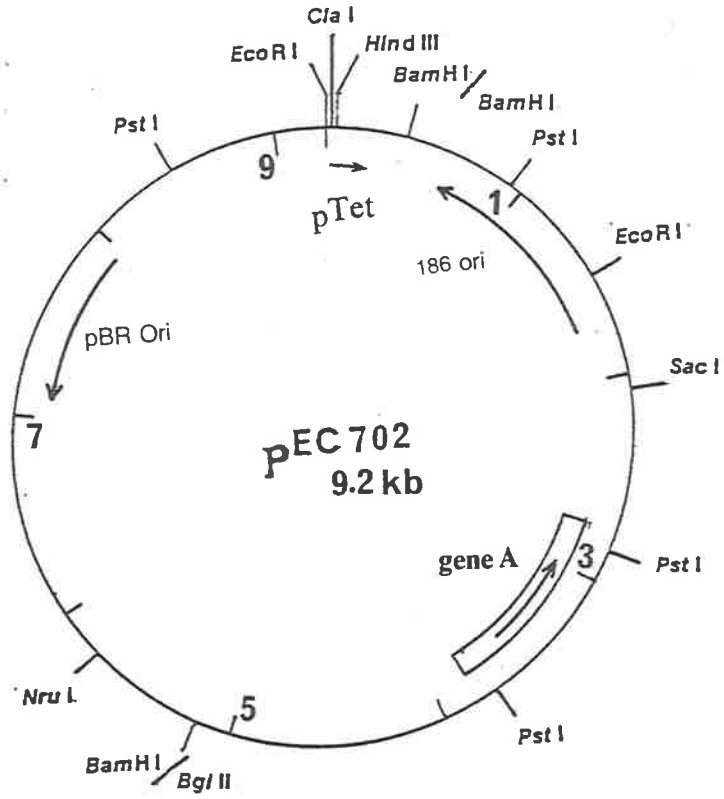

1
2
3
2 3.1
9.4
6.5
4.4
2.3
2.0
7.6
5.2
4.0
1.6

Fig. 3.4 Circular maps of pEC701 and pEC702


Cla I
EcoR I
Hind III
BamH I
Bgl II
Pst I
9
pTet
1
pBR Ori
Pst I
7
pEC 701
9.2 kb
gene A
Pst I
3
Nru I
186 ori
5
Sac I
BamH I
BamH I
Pst I
EcoR I


Cla I
EcoR I
Hind III
BamH I
BamH I
Pst I
Pst I
9
pTet
1
EcoR I
186 ori
pBR Ori
Sac I
7
pEC 702
9.2 kb
gene A
3
Pst I
Nru I
5
Pst I
BamH I
Bgl II

carrying pEC701 was later found to be due to the phage-specific replication of this plasmid.

3.3.3 FUNCTIONAL CHARACTERIZATION OF pEC701

Since the orientation of insert in pEC701 was such that p^{tet} would transcribe the BGLBAM fragment rightwards, the direction in which the early region of phage DNA was transcribed *in vivo* (B. Kalionis, personal communication), one would expect that *geneA* might be expressed on this plasmid. As it was not known whether a phage promoter was present on the BGLBAM fragment, or an anti-clockwise promoter was present to the right of *BamHI* site on pBR322, expression of *geneA* in pEC702 could not be predicted. Determining the expression of *geneA* in pEC701 and pEC702 by its ability to replicate these plasmids from 186*ori* would resolve this at the same time it confirmed that BGLBAM fragment contained all phage replication functions.

3.3.3.1 Evidence for the phage-specific replication of pEC701

As ColEI replication is inhibited in DNA polymeraseI deficient (*polA1*) strain (Delucia and Cairns, 1969; Kornberg, 1980; Cooper and Hanawalt, 1972; Kingsbury and Helinski, 1973; Scott, 1984) ability of pEC701 and pEC702 to transform a *polA* strain will indicate the presence of an alternative function origin in these plasmids. Phage-specific replication of these plasmids was therefore tested by transforming a *polA* strain, (E0941) and it was found that while pEC701 could confer amp^R character to the cells, neither pEC702 nor pBR322 could transform the *polA* cells. This indicated that the 186*ori* carried on the former was active but the *ori* carried on pEC702

was not, presumably due to the lack of transcription of *geneA*. The inability of pEC703 (= pEC701*Aam24*), pEC704 (= pEC701-*Aam43*) and pEC450 (= pEC701*Aam11*; S. Williams; personal communication) to transform *polA* strain confirmed that the product of *geneA* was necessary for 186-specific replication of the plasmid. Presence of the *rep* mutation, which inhibits 186 replication (Calendar, *et.al.*, 1970), together with *polA* in E0944 resulted in the inability of pEC701 to transform this strain.

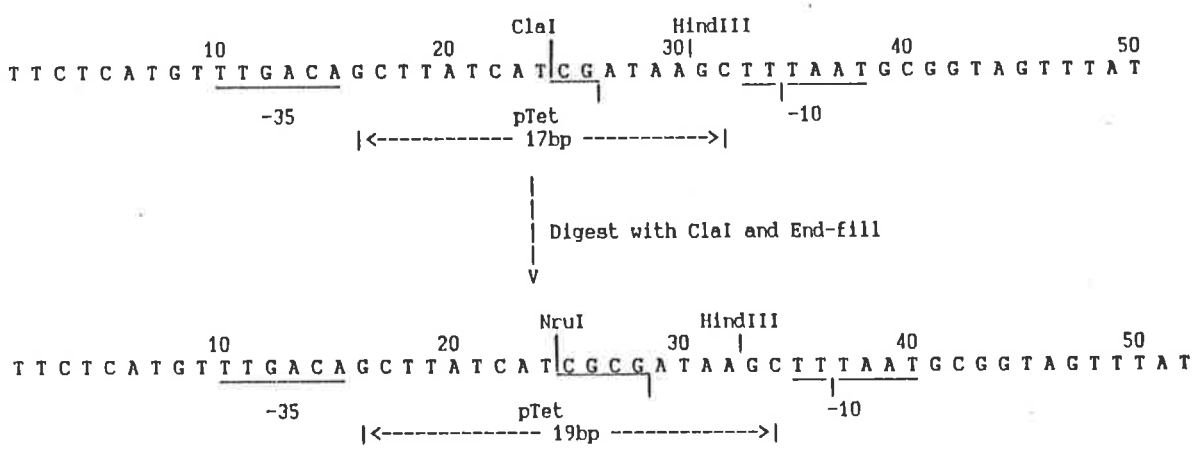
3.3.3.2 Transcription from p^{tet} is needed for the phage-specific replication.

Since the absence of 186-specific replication in pEC702 seemed to be due to the orientation of the insert relative to p^{tet}, the phage-specific replication of pEC701 must be dependent on transcription from this promoter. To test if inactivation of p^{tet} resulted in loss of 186-specific replication of pEC701 the unique *ClaI* site of this plasmid was mutated by end-filling. As the *ClaI* site is located between the -35 and -10 regions of p^{tet} (Fig. 3.5; Siebenlist *et.al.*, 1980; Sutcliffe, 1979), end-filling will increase the spacing between these regions and consequently reduce or abolish the activity of the promoter. It was found that such altered plasmids were unable to transform *polA*- strains which proved that the transcription from p^{tet} was essential for the initiation of replication from 186*ori* present on pEC701. This incidentally supported the earlier results of *in vivo* transcription studies which had indicated that the major lytic

Fig. 3.5 Sequence of pBR322 showing p^{tet}

The *ClaI* site, located at position 23 of pBR322 (Sutcliffe, 1979), when end-filled will generate *NruI* site by adding the dinucleotide CG.

The spacing between the -35 and -10 regions of p^{tet} are shown below the sequence.


ClaI
HindIII
10
20
30
40
50
T T C T C A T G T T T G A C A G C T T A T C A T C G A T A A G C T T T A A T G C G G T A G T T T A T
-35
pTet
-10
17bp
Digest with ClaI and End-fill
NruI
HindIII
10
20
30
40
50
T T C T C A T G T T T G A C A G C T T A T C A T C G C G A T A A G C T T T A A T G C G G T A G T T T A T
-35
pTet
-10
19bp

promoter of 186 was situated at 74% of the chromosome (B. Kalionis, personal communication).

3.3.3.3 Phage-specific replication of pEC701 produced minute colonies.

Characteristically the strains carrying pEC701 produced minute colonies in both *polA+* and *polA* background (Fig. 3.6). This minute colony character seemed to be associated with the replication of 186*ori* as pEC701 variants which did not transform *polA* strain produced normal sized colonies when transforming *polA+* strain.

The plasmid pEC701 should give normal sized colonies in a *rep-* background if the minute colony character was associated with the replication of phage *ori*. This was in fact found to be true when the plasmid was introduced into a *rep3* strain (E4011).

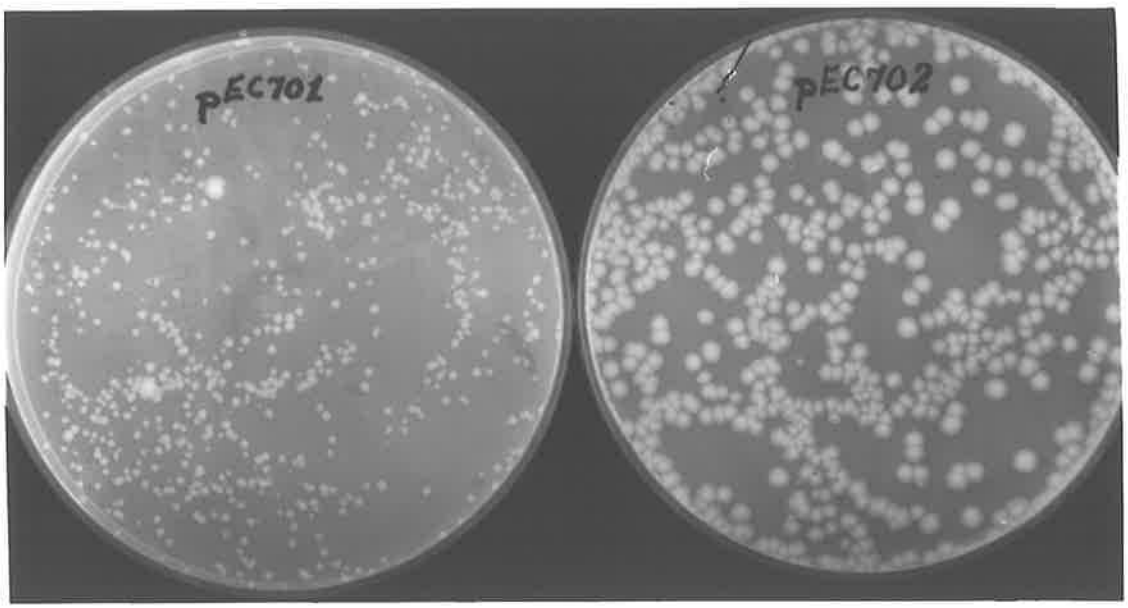
The minute colony character appeared to be due to the slow growth rate of cells as the generation time of the strain carrying pEC701 (E2249) was about 3 hours at 37C compared with the generation time of 30 minutes for the strain E2250 which carried pEC702 (E2250).

3.3.3.4 Colony size revertants

A high rate of reversion to normal sized colonies was apparent when overnight culture of the strain carrying pEC701 (E2249) was plated for single colonies. The normal sized colonies appeared to carry the original plasmid, as indicated by the fact that the plasmid DNA isolated from them gave

Fig 3.6 Colony sizes of E2249 and E2250

Single colonies of E2249 (= E536(pEC701)) and E2250 (= E536(pEC702)), resuspended in LB, were plated on YGC + Amp (50 ng/ml) and incubated at 37°C for 36 hours. These were, then, photographed under short wave UV.


pEC701
pEC702

minute colonies on re-transformation of the parental strain, E536. Presumably the host chromosome in the normal colony variants carried mutation(s) which prevented 186-specific replication of pEC701. This was supported by the finding that curing these cells of the plasmid did not change their 186R character (S. Williams; personal communication).

Normal sized colonies could also be obtained from freshly transformed cell population, and in this case the plasmid DNA from such colonies on re-transformation of E536 gave only normal sized colonies which indicated that the mutation(s) responsible for the loss of 186-specific replication of pEC701 was carried on the plasmid itself. As expected, these colonies were sensitive to 186 infection even though the plasmid DNA still resided within the cells.

3.3.3.5 The Yield of pEC701 is high from Stationary Phase Cultures.

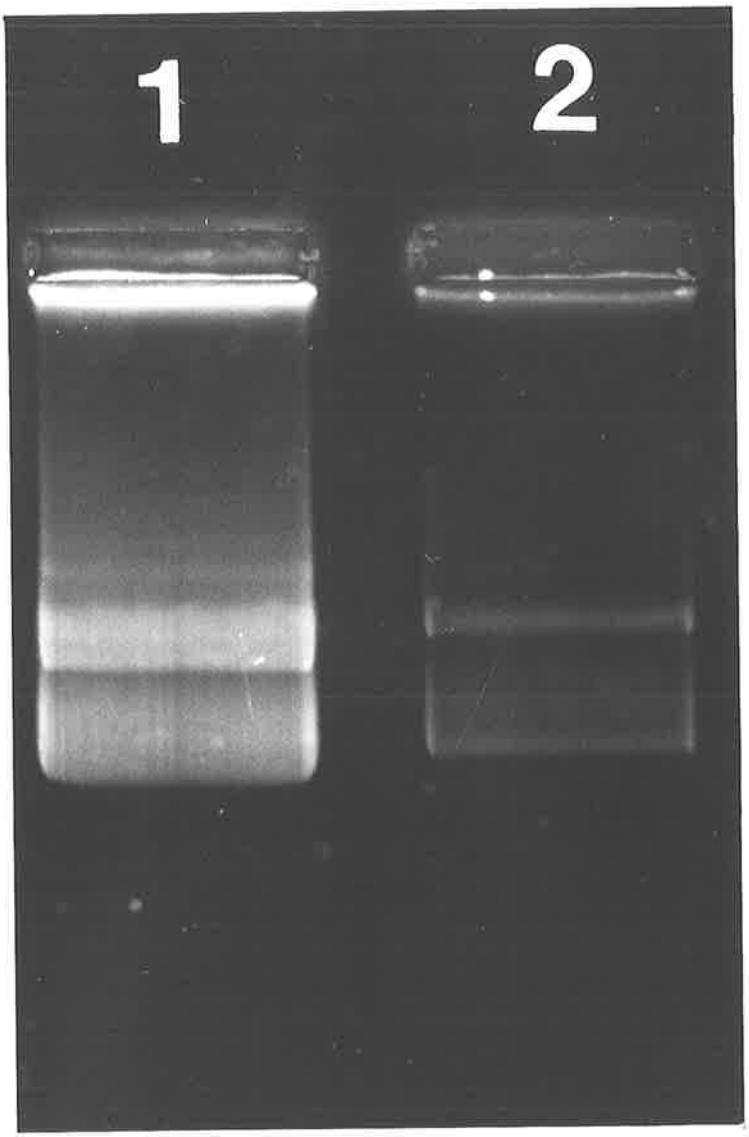
Comparison of plasmid preparations from strains E2249 and E2250 showed that the yield of pEC701 DNA was significantly higher than that of pEC702 (Fig. 3.7). This high yield was not seen when replication of *ori* was blocked by the presence of *Aamber* mutations. The increased yield of pEC701 probably reflected the high copy number of the plasmid. Interestingly this high yield was obtained only when the cultures reached stationary phase. Exponentially growing cultures of E2249 yielded lesser plasmid DNA than those of E2250.

Fig 3.7 Comparison of the yields of pEC701 and pEC702

Overnight cultures of strains E2249 (= E536(pEC701)) and E2250 (= E536(pEC702)) were prepared in LB + Amp (50 ng/ml) at 37°C. Plasmid DNA was prepared for 1 ml each of these cultures, treated with *RNAseA* and the entire amount was electrophoresed on 1% agarose gel.

Tracks:

1. Plasmid DNA from E2249
2. Plasmid DNA from E2250


1
2

DISCUSSION

The results presented in this chapter proved that the BGLBAM fragment carried all phage functions required for 186 replication, as shown by the ability of pEC701 to replicate from the phage *ori*. The phage-specific replication of pEC701 was dependent on the expression of *geneA*, which in turn depended on the transcription initiated from p^{tet}. The host mutation *rep3*, which blocks the replication of phage 186, was found to block the initiation of replication from the phage *ori* on pEC701.

Phage-specific replication of pEC701 was associated with minute colony character which incidentally provided an easily identifiable phenotypic character to assess the presence of functional 186*ori* on this plasmid. The minute colony character apparently resulted from the slow growth rate of the strains carrying the plasmid pEC701.

Spontaneous host and plasmid mutations which prevented replication of the phage *ori* resulted in normal colony size revertants and could be potential source material for the identification of the host and phage functions required for the replication of phage 186.

Since the BGLBAM fragment encoded all phage functions essential for 186 replication, DNA sequencing of this fragment might help to identify the genes and the structural features essential for the replication of this phage.

CHAPTER 4

SEQUENCE OF THE BGLBAM FRAGMENT

4.1 INTRODUCTION

The ability of pEC701 to replicate from 186 *ori* suggested that all phage functions essential for replication were carried on the BGLBAM fragment. Determining the DNA sequence of this fragment would help to identify the potential replication genes. The sequence might also reveal sites of restriction enzymes which could be used to prune down the BGLBAM fragment to define the replication genes as well as to identify the minimal DNA required for replication.

The *ori* region of lambdoid phages show several potential secondary structures which are involved in the binding of replication proteins (Gromchedl and Hobom, 1979; Hobom and Lusky, 1980; Hobom, 1981). Presence of such structures in 186*ori*, if any, can be predicted from the DNA sequence and may be of help in understanding the DNA-protein interactions required for the initiation of replication. Identification of the exact site of initiation of DNA synthesis would also be facilitated if the DNA sequence was available.

This chapter describes the DNA sequence determination of the BGLBAM fragment and the analysis of the sequence to expose the potential replication genes.

4.2 DETERMINATION OF THE SEQUENCE

4.2.1 Methology

A modified version of the dideoxy-chain termiation method of Sanger *et.al.* (1977, 1980) was used to determine the DNA sequence, as described in Chapter 2. Sub-cloning of

restriction fragments produced by digestion with four-base restriction enzymes was preferred to sonication of DNA (Fuhrman *et.al.* 1981) mainly because this would help to build a library of clones which were easily reproducible. This was considered advantageous as it was anticipated that cloning of specific fragments carrying various mutations which blocked replication might become necessary at later stages. In such instances wild-type sequence for isolation and comparison of the clones could be obtained from the library.

4.2.2 Strategy for Sub-cloning and sequencing of the BGLBAM fragment

The BGLBAM fragment was sequenced in two stages. Firstly the *SacI* to *BamHI* DNA fragment (90 to 96% of the 186 chromosome) was isolated from pEC701 and sequenced after cloning the sub-digestion fragments into M13mp7. Fragments generated with the four-base restriction enzymes *HaeIII*, *HpaII*, and *TaqI* were used in this instance.

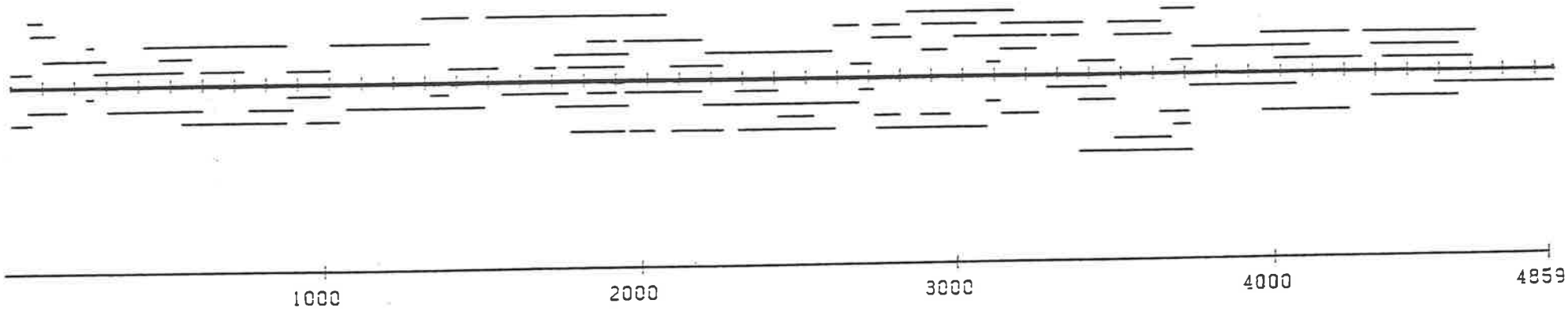
In the second stage of sub-cloning the *EcoRI* to *EcoRI* fragment of pEC701, which encodes the 79.6 to 92% of 186 chromosome together with 375 bases of pBR322 DNA (*EcoRI* to *BamHI*), was sub-digested separately with *AluI*, *FnuDII*, *HhaI* and *HpaII*, and sequenced after shot-gun cloning into M13mp9. This sequence therefore overlapped with the *SacI-BamHI* sequence for 613 bases.

4.2.3 Sequencing

Sequencing was generally done by using the universal primer (17-mer; 5'-GTAAAACGACGGCCAGT-3) on clones selected at

Fig. 4.1 Strategy of sequencing the BGLBAM fragment

Horizontal bars correspond in length to the actual number of bases read off individual sequencing experiments. Sequences read off the *l*-strand are given above the solid line and those of the *r*-strand are given below this line. In most cases one end of the bar marks the restriction site used for genearting the clone.


1000
2000
3000
4000
4859

random. At later stages of sequencing specific primers were used on larger clones in order to fill the gaps in the sequence. The specific primers were prepared by isolating the appropriate restriction fragments from pEC701 DNA. Sections of the sequence which were read only in one direction in the completed sequence were sequenced again in the opposite direction by reversing the clones. The strategy map of the completed sequence is given in Figure 4.1. The number of times a base pair was sequenced varied between 3 and 11 in either direction. No disparity was observed between individual bases read off different clones but disparities due to "compressions" were seen in some instances and in such areas the sequence of the opposite strand was taken as the correct one. The completed sequence is given in Figure 4.2 and contains 4859 base pairs.

4.2.4 Accuracy of the sequence

Since most areas had been sequenced from more than one clone and in both directions, no serious frameshift or base substitution error was anticipated. Absence of frameshift errors in the coding regions of genes was also confirmed by the codon preference plot (see Fig. 4.3).

As a further check on the accuracy of ordering of the restriction fragments in the completed sequence, the plasmid pEC701 DNA was digested with *HinfI* and the size and distributions of bands were compared with the pattern predicted from the sequence (Fig. 4.4). The actual distribution of bands agreed exactly with the predicted pattern thereby confirming the accuracy of ordering of the fragments.

Fig. 4.2 Sequence of the 79.6 to 96.0% of the 186 chromosome (the BGLBAM fragment)

The upper strand (*l*-strand) is written from 5'-3'. Important regulatory sequences and restriction sites are underlined.

Amino acid sequence of the potential genes are shown above the DNA sequence.

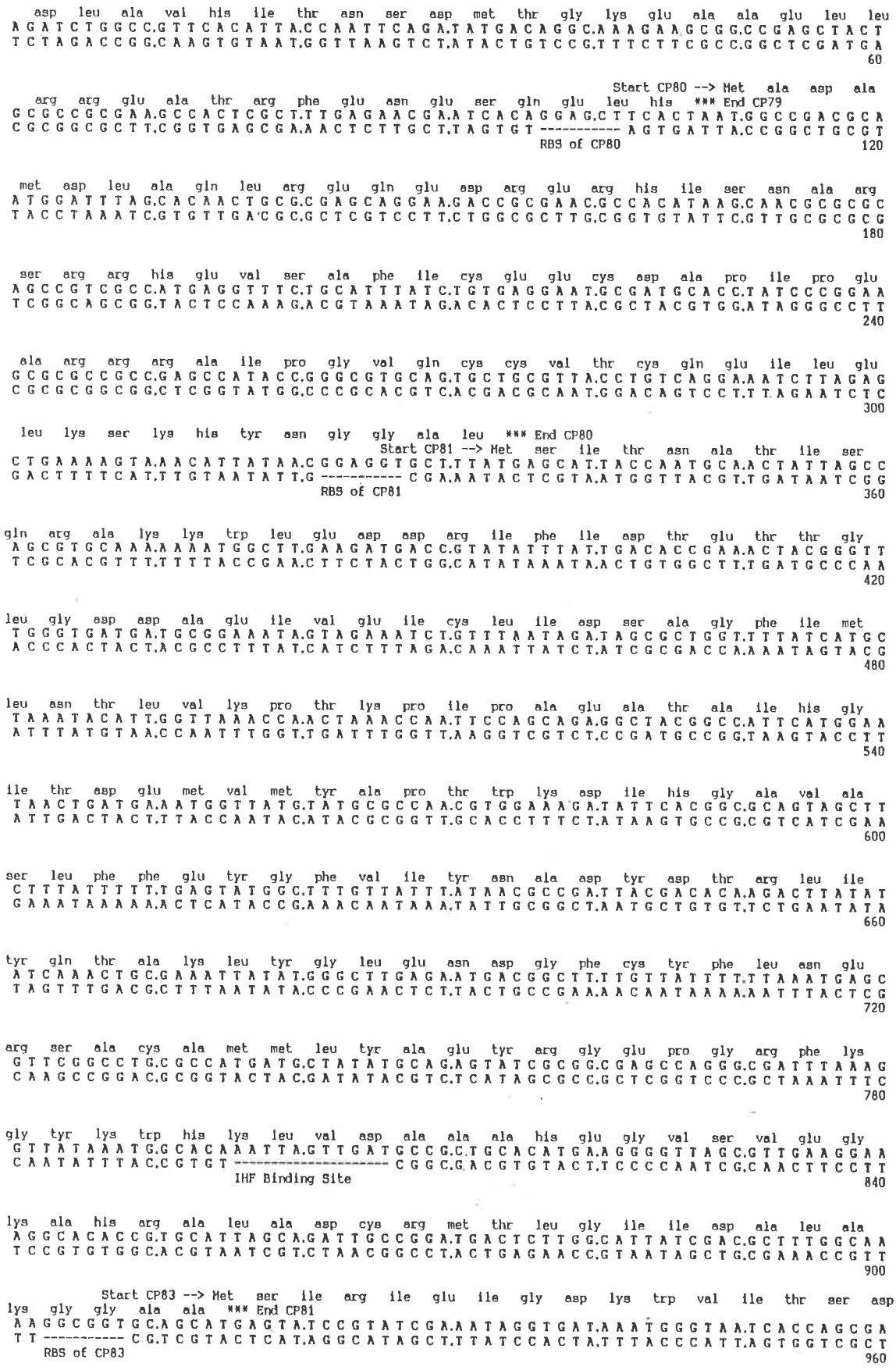
The potential gene *CP79* starts from an ATG codon at 125 bases to the left of the *BglII* site (H. Richardson, personal communication).

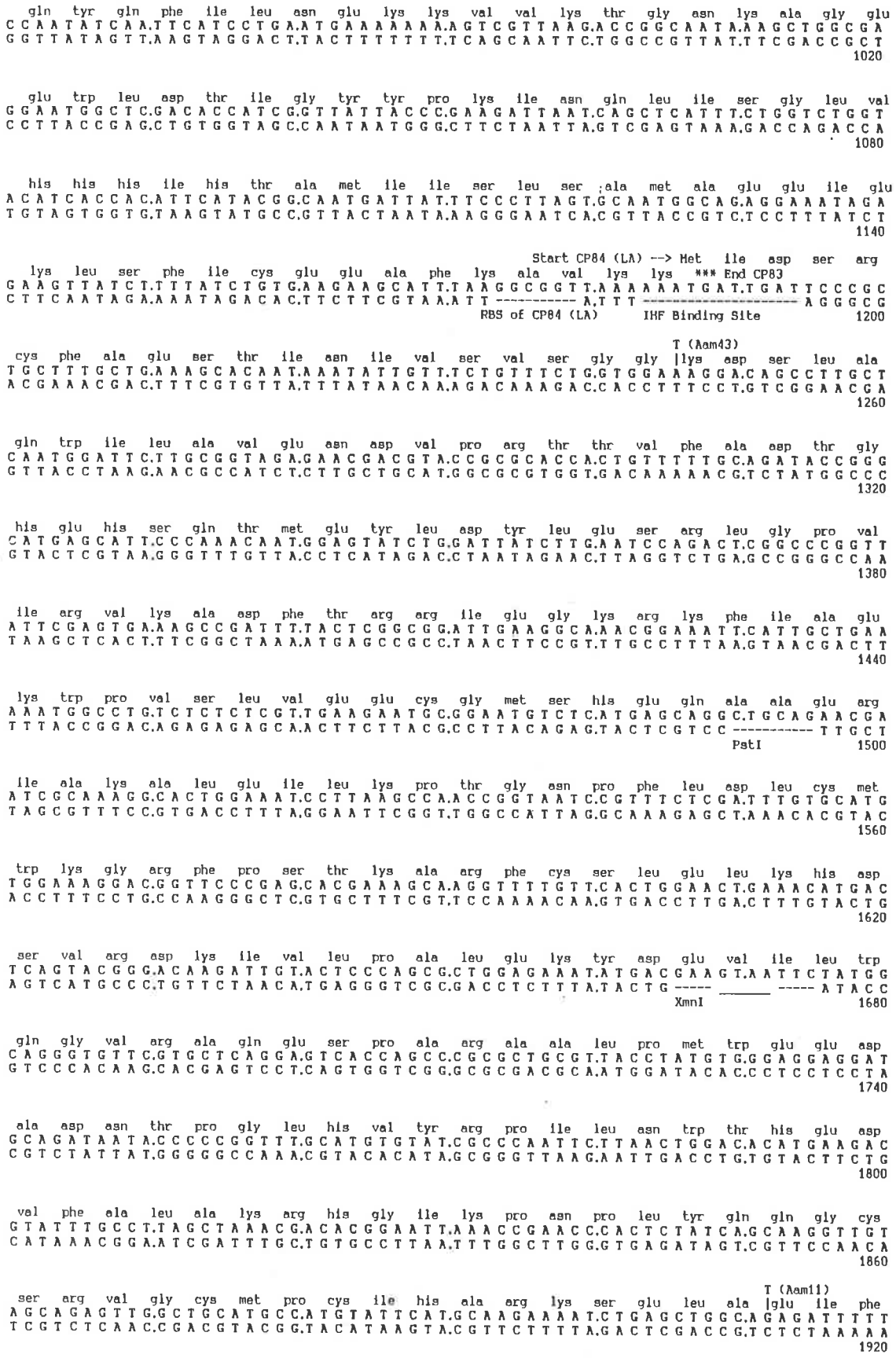
Consensus sequence for the *lexA* binding site as given by Walker (1984) is: 5' t a C T G T a t a t a - a - a C A G t a 3'.

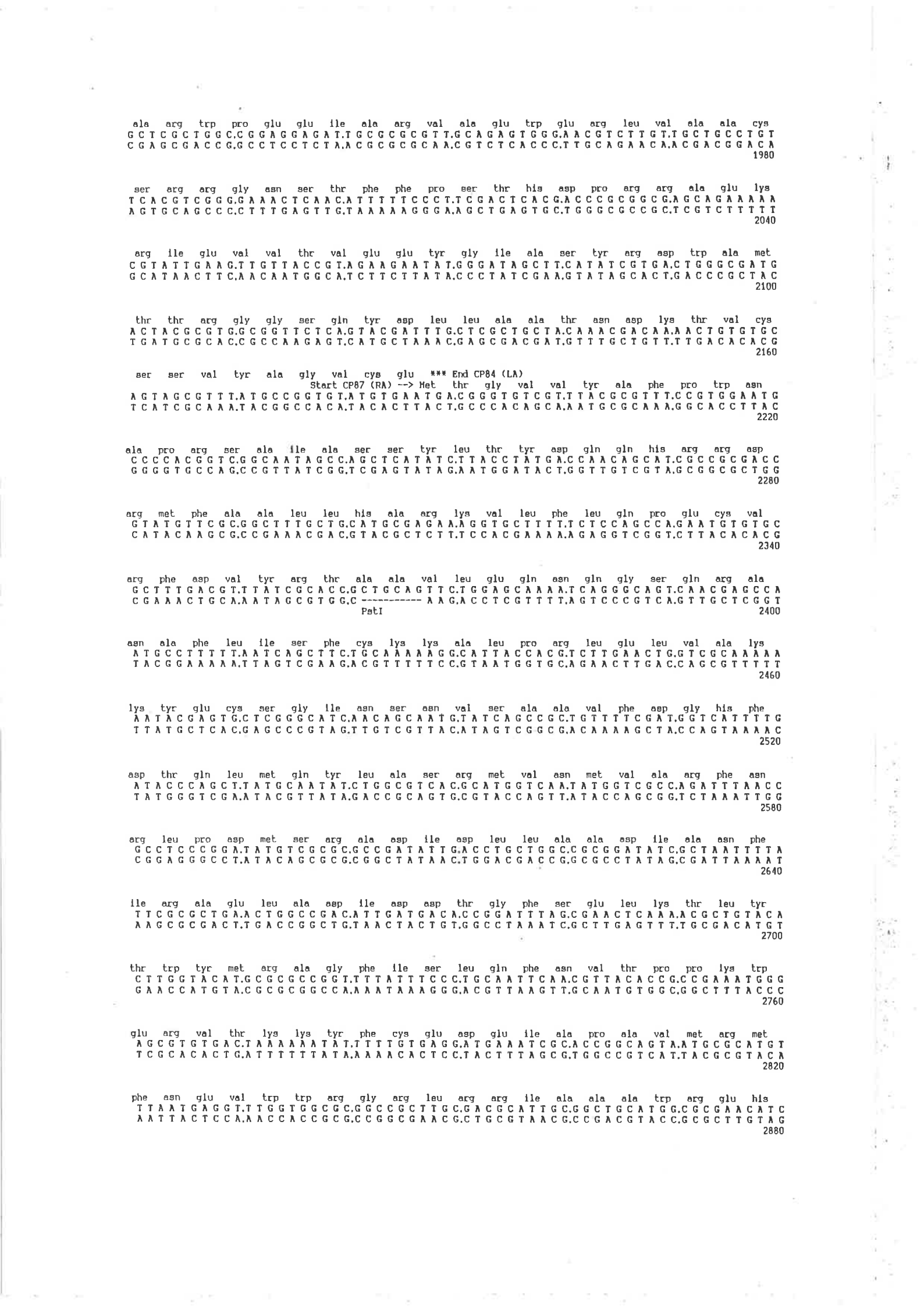
The sequence for the *dnaA* binding site as given by Ohmori *et. al.* (1984) is: 5' T G T G G A T A G A 3'.

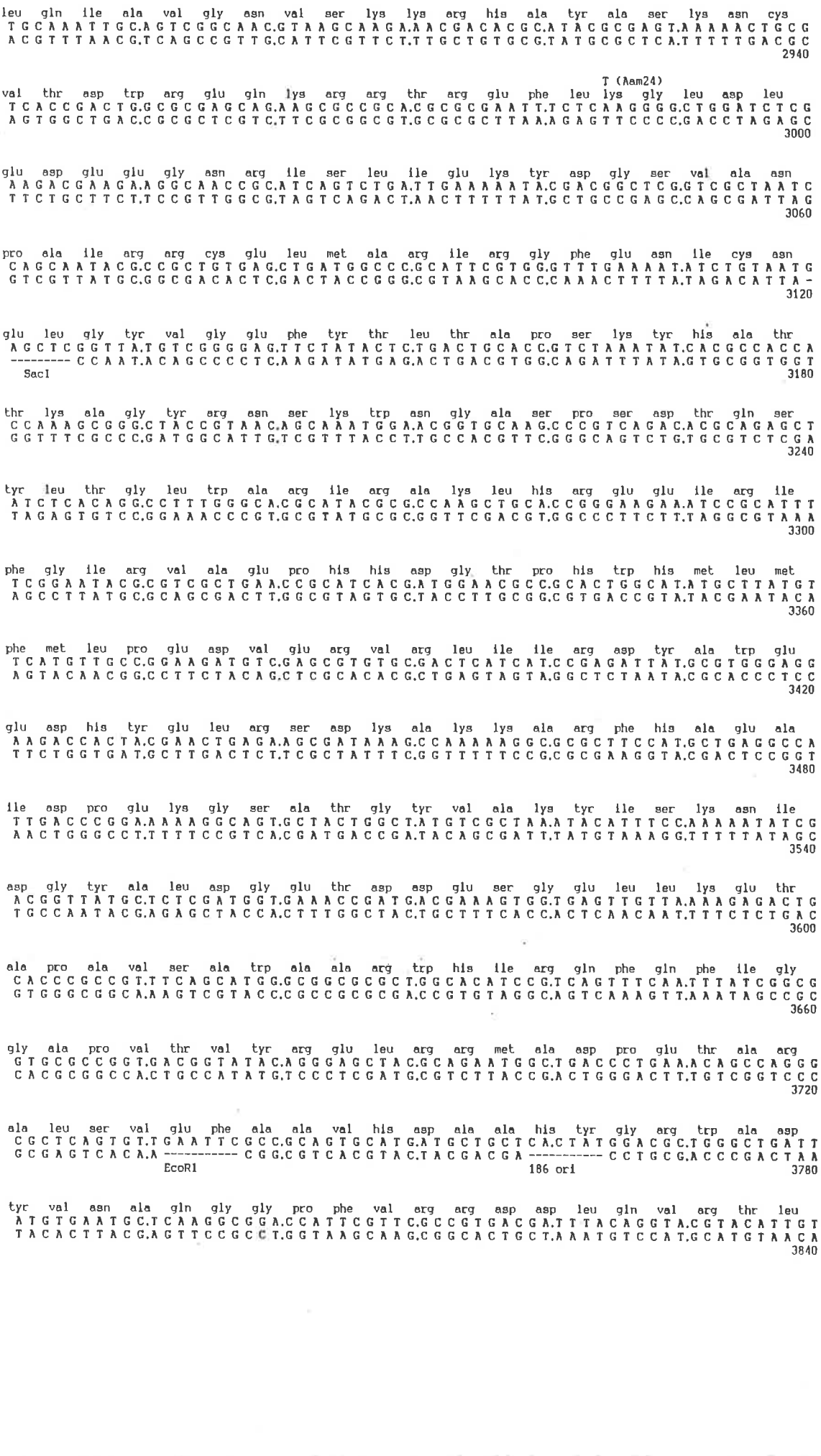
Consensus sequence for the IHF binding site as given by Craig and Nash (Unpublished) is: 5' A A - - - Pu T T G A T 3'.

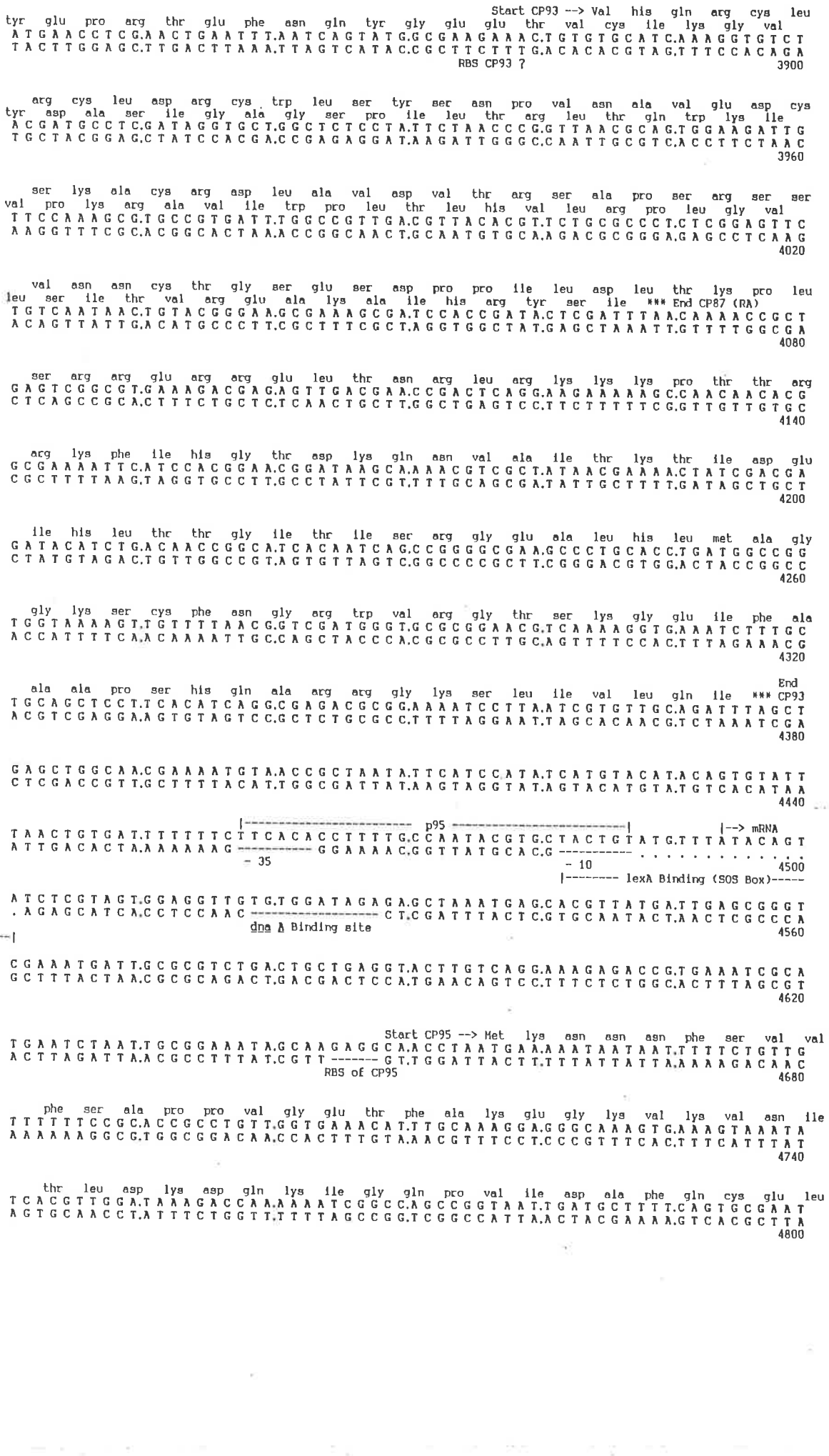
Ribosome binding site as given by Shine and Dalgarno (1974) is: 5' G G A G G T 3'.


asp leu ala val his ile thr asn ser asp met thr gly lys glu ala ala glu leu leu
A G A T C T G G C C.G T T C A C A T T A.C C A A T T C A G A.T A T G A C A G G C.A A A G A A G C G G.C C G A G C T A C T
T C T A G A C C G G.C A A G T G T A A T.G G T T A A G T C T.A T A C T G T C C G.T T T C T T C G C C.G G C T C G A T G A
60
Start CP80 --> Met ala asp ala
arg arg glu ala thr arg phe glu asn glu ser gln glu leu his *** End CP79
G C G C C G C G A A.G C C A C T C G C T.T T G A G A A C G A.A T C A C A G G A G.C T T C A C T A A T.G G C C G A C G C A
C G C G G C G C T T.C G G T G A G C G A.A A C T C T T G C T.T A G T G T ---------- A G T G A T T A.C C G G C T G C G T
RBS of CP80
120
met asp leu ala gln leu arg glu gln glu asp arg glu arg his ile ser asn ala arg
A T G G A T T T A G.C A C A A C T G C G.C G A G C A G G A A.G A C C G C G A A C.G C C A C A T A A G.C A A C G C G C G C
T A C C T A A A T C.G T G T T G A C G C.G C T C G T C C T T.C T G G C G C T T G.C G G T G T A T T C.G T T G C G C G C G
180
ser arg arg his glu val ser ala phe ile cys glu glu cys asp ala pro ile pro glu
A G C C G T C G C C.A T G A G G T T T C.T G C A T T T A T C.T G T G A G G A A T.G C G A T G C A C C.T A T C C C G G A A
T C G G C A G C G G.T A C T C C A A A G.A C G T A A A T A G.A C A C T C C T T A.C G C T A C G T G G.A T A G G G C C T T
240
ala arg arg arg ala ile pro gly val gln cys cys val thr cys gln glu ile leu glu
G C G C G C C G C C.G A G C C A T A C C.G G G C G T G C A G.T G C T G C G T T A.C C T G T C A G G A.A A T C T T A G A G
C G C G C G G C G G.C T C G G T A T G G.C C C G C A C G T C.A C G A C G C A A T.G G A C A G T C C T.T T A G A A T C T C
300
leu lys ser lys his tyr asn gly gly ala leu *** End CP80
Start CP81 --> Met ser ile thr asn ala thr ile ser
C T G A A A A G T A.A A C A T T A T A A.C G G A G G T G C T.T T A T G A G C A T.T A C C A A T G C A.A C T A T T A G C C
G A C T T T T C A T.T T G T A A T A T T.G ---------- C G A.A A T A C T C G T A.A T G G T T A C G T.T G A T A A T C G G
RBS of CP81
360
gln arg ala lys lys trp leu glu asp asp arg ile phe ile asp thr glu thr thr gly
A G C G T G C A A A.A A A A T G G C T T.G A A G A T G A C C.G T A T A T T T A T.T G A C A C C G A A.A C T A C G G G T T
T C G C A C G T T T.T T T T A C C G A A.C T T C T A C T G G.C A T A T A A A T A.A C T G T G G C T T.T G A T G C C C A A
420
leu gly asp asp ala glu ile val glu ile cys leu ile asp ser ala gly phe ile met
T G G G T G A T G A.T G C G G A A A T A.G T A G A A A T C T.G T T T A A T A G A.T A G C G C T G G T.T T T A T C A T G C
A C C C A C T A C T.A C G C C T T T A T.C A T C T T T A G A.C A A A T T A T C T.A T C G C G A C C A.A A A T A G T A C G
480
leu asn thr leu val lys pro thr lys pro ile pro ala glu ala thr ala ile his gly
T A A A T A C A T T.G G T T A A A C C A.A C T A A A C C A A.T T C C A G C A G A.G G C T A C G G C C.A T T C A T G G A A
A T T T A T G T A A.C C A A T T T G G T.T G A T T T G G T T.A A G G T C G T C T.C C G A T G C C G G.T A A G T A C C T T
540
ile thr asp glu met val met tyr ala pro thr trp lys asp ile his gly ala val ala
T A A C T G A T G A.A A T G G T T A T G.T A T G C G C C A A.C G T G G A A A G A.T A T T C A C G G C.G C A G T A G C T T
A T T G A C T A C T.T T A C C A A T A C.A T A C G C G G T T.G C A C C T T T C T.A T A A G T G C C G.C G T C A T C G A A
600
ser leu phe phe glu tyr gly phe val ile tyr asn ala asp tyr asp thr arg leu ile
C T T T A T T T T T.T G A G T A T G G C.T T T G T T A T T T.A T A A C G C C G A.T T A C G A C A C A.A G A C T T A T A T
G A A A T A A A A A.A C T C A T A C C G.A A A C A A T A A A.T A T T G C G G C T.A A T G C T G T G T.T C T G A A T A T A
660
tyr gln thr ala lys leu tyr gly leu glu asn asp gly phe cys tyr phe leu asn glu
A T C A A A C T G C.G A A A T T A T A T.G G G C T T G A G A.A T G A C G G C T T.T G T T A T T T T.T T A A A T G A G C
T A G T T T G A C G.C T T T A A T A T A.C C C G A A C T C T.T A C T G C C G A A.A A C A A T A A A A.A A T T T A C T C G
720
arg ser ala cys ala met met leu tyr ala glu tyr arg gly glu pro gly arg phe lys
G T T C G G C C T G.C G C C A T G A T G.C T A T A T G C A G.A G T A T C G C G G.C G A G C C A G G G.C G A T T T A A A G
C A A G C C G G A C.G C G G T A C T A C.G A T A T A C G T C.T C A T A G C G C C.G C T C G G T C C C.G C T A A A T T T C
780
gly tyr lys trp his lys leu val asp ala ala ala his glu gly val ser val glu gly
G T T A T A A A T G.G C A C A A A T T A.G T T G A T G C C G.C T G C A C A T G A.A G G G G T T A G C.G T T G A A G G A A
C A A T A T T T A C.C G T G T ---------- C G G C.G A C G T G T A C T.T C C C C A A T C G.C A A C T T C C T T
IHF Binding Site
840
lys ala his arg ala leu ala asp cys arg met thr leu gly ile ile asp ala leu ala
A G G C A C A C C G.T G C A T T A G C A.G A T T G C C G G A.T G A C T C T T G G.C A T T A T C G A C.G C T T T G G C A A
T C C G T G T G G C.A C G T A A T C G T.C T A A C G G C C T.A C T G A G A A C C.G T A A T A G C T G.C G A A A C C G T T
900
Start CP83 --> Met ser ile arg ile glu ile gly asp lys trp val ile thr ser asp
lys gly gly ala ala *** End CP81
A A G G C G G T G C.A G C A T G A G T A.T C C G T A T C G A.A A T A G G T G A T.A A A T G G G T A A.T C A C C A G C G A
T T ---------- C G.T C G T A C T C A T.A G G C A T A G C T.T T A T C C A C T A.T T T A C C C A T T.A G T G G T C G C T
RBS of CP83
960


gln tyr gln phe ile leu asn glu lys lys val val lys thr gly asn lys ala gly glu
1020
glu trp leu asp thr ile gly tyr tyr pro lys ile asn gln leu ile ser gly leu val
1080
his his his ile his thr ala met ile ile ser leu ser ;ala met ala glu glu ile glu
1140
Start CP84 (LA) --> Met ile asp ser arg
lys leu ser phe ile cys glu glu ala phe lys ala val lys lys *** End CP83
RBS of CP84 (LA)
IHF Binding Site
1200
T (Aam43)
cys phe ala glu ser thr ile asn ile val ser val ser gly gly lys asp ser leu ala
1260
gln trp ile leu ala val glu asn asp val pro arg thr thr val phe ala asp thr gly
1320
his glu his ser gln thr met glu tyr leu asp tyr leu glu ser arg leu gly pro val
1380
ile arg val lys ala asp phe thr arg arg ile glu gly lys arg lys phe ile ala glu
1440
lys trp pro val ser leu val glu glu cys gly met ser his glu gln ala ala glu arg
PstI
1500
ile ala lys ala leu glu ile leu lys pro thr gly asn pro phe leu asp leu cys met
1560
trp lys gly arg phe pro ser thr lys ala arg phe cys ser leu glu leu lys his asp
1620
ser val arg asp lys ile val leu pro ala leu glu lys tyr asp glu val ile leu trp
XmnI
1680
gln gly val arg ala gln glu ser pro ala arg ala ala leu pro met trp glu glu asp
1740
ala asp asn thr pro gly leu his val tyr arg pro ile leu asn trp thr his glu asp
1800
val phe ala leu ala lys arg his gly ile lys pro asn pro leu tyr gln gln gly cys
1860
T (Aam11)
ser arg val gly cys met pro cys ile his ala arg lys ser glu leu ala glu ile phe
1920


ala arg trp pro glu glu ile ala arg val ala glu trp glu arg leu val ala ala cys
1980
ser arg arg gly asn ser thr phe phe pro ser thr his asp pro arg arg ala glu lys
2040
arg ile glu val val thr val glu glu tyr gly ile ala ser tyr arg asp trp ala met
2100
thr thr arg gly gly ser gln tyr asp leu leu ala ala thr asn asp lys thr val cys
2160
ser ser val tyr ala gly val cys glu *** End CP84 (LA)
Start CP87 (RA) --> Met thr gly val val tyr ala phe pro trp asn
2220
ala pro arg ser ala ile ala ser ser tyr leu thr tyr asp gln gln his arg arg asp
2280
arg met phe ala ala leu leu his ala arg lys val leu phe leu gln pro glu cys val
2340
arg phe asp val tyr arg thr ala ala val leu glu gln asn gln gly ser gln arg ala
PstI
2400
asn ala phe leu ile ser phe cys lys lys ala leu pro arg leu glu leu val ala lys
2460
lys tyr glu cys ser gly ile asn ser asn val ser ala ala val phe asp gly his phe
2520
asp thr gln leu met gln tyr leu ala ser arg met val asn met val ala arg phe asn
2580
arg leu pro asp met ser arg ala asp ile asp leu leu ala ala asp ile ala asn phe
2640
ile arg ala glu leu ala asp ile asp asp thr gly phe ser glu leu lys thr leu tyr
2700
thr trp tyr met arg ala gly phe ile ser leu gln phe asn val thr pro pro lys trp
2760
glu arg val thr lys lys tyr phe cys glu asp glu ile ala pro ala val met arg met
2820
phe asn glu val trp trp arg gly arg leu arg arg ile ala ala ala trp arg glu his
2880


leu gln ile ala val gly asn val ser lys lys arg his ala tyr ala ser lys asn cys
T (Aam24)
val thr asp trp arg glu gln lys arg arg thr arg glu phe leu lys gly leu asp leu
glu asp glu glu gly asn arg ile ser leu ile glu lys tyr asp gly ser val ala asn
pro ala ile arg arg cys glu leu met ala arg ile arg gly phe glu asn ile cys asn
glu leu gly tyr val gly glu phe tyr thr leu thr ala pro ser lys tyr his ala thr
SacI
thr lys ala gly tyr arg asn ser lys trp asn gly ala ser pro ser asp thr gln ser
tyr leu thr gly leu trp ala arg ile arg ala lys leu his arg glu glu ile arg ile
phe gly ile arg val ala glu pro his his asp gly thr pro his trp his met leu met
phe met leu pro glu asp val glu arg val arg leu ile ile arg asp tyr ala trp glu
glu asp his tyr glu leu arg ser asp lys ala lys lys ala arg phe his ala glu ala
ile asp pro glu lys gly ser ala thr gly tyr val ala lys tyr ile ser lys asn ile
asp gly tyr ala leu asp gly glu thr asp asp glu ser gly glu leu leu lys glu thr
ala pro ala val ser ala trp ala ala arg trp his ile arg gln phe gln phe ile gly
gly ala pro val thr val tyr arg glu leu arg arg met ala asp pro glu thr ala arg
ala leu ser val glu phe ala ala val his asp ala ala his tyr gly arg trp ala asp
EcoR1
186 ori
tyr val asn ala gln gly gly pro phe val arg arg asp asp leu gln val arg thr leu


Start CP93 --> Val his gln arg cys leu
tyr glu pro arg thr glu phe asn gln tyr gly glu glu thr val cys ile lys gly val
RBS CP93 ?
3900
arg cys leu asp arg cys trp leu ser tyr ser asn pro val asn ala val glu asp cys
tyr asp ala ser ile gly ala gly ser pro ile leu thr arg leu thr gln trp lys ile
3960
ser lys ala cys arg asp leu ala val asp val thr arg ser ala pro ser arg ser ser
val pro lys arg ala val ile trp pro leu thr leu his val leu arg pro leu gly val
4020
val asn asn cys thr gly ser glu ser asp pro pro ile leu asp leu thr lys pro leu
leu ser ile thr val arg glu ala lys ala ile his arg tyr ser ile *** End CP87 (RA)
4080
ser arg arg glu arg arg glu leu thr asn arg leu arg lys lys lys pro thr thr arg
4140
arg lys phe ile his gly thr asp lys gln asn val ala ile thr lys thr ile asp glu
4200
ile his leu thr thr gly ile thr ile ser arg gly glu ala leu his leu met ala gly
4260
gly lys ser cys phe asn gly arg trp val arg gly thr ser lys gly glu ile phe ala
4320
End
ala ala pro ser his gln ala arg arg gly lys ser leu ile val leu gln ile *** CP93
4380
4440
p95
--> mRNA
- 35
- 10
lexA Binding (SOS Box)
4500
dnaA Binding site
4560
4620
Start CP95 --> Met lys asn asn asn phe ser val val
RBS of CP95
4680
phe ser ala pro pro val gly glu thr phe ala lys glu gly lys val lys val asn ile
4740
thr leu asp lys asp gln lys ile gly gln pro val ile asp ala phe gln cys glu leu
4800

```
     thr  lys  arg  ile  gln  ser  val  phe  pro  ser  thr  arg  val  thr  val  lys  lys  gly  ser
T G A C C A A G C G.A A T A C A G T C C.G T T T T C C C G T.C A A C G C G C G T.T A C T G T T A A A.A A G G G A T C C
A C T G G T T C G C.T T A T G T C A G G.C A A A A G G G C A.G T T G C G C G C A.A T G A C A A T T T.T T C ----------
                                                                                                    BamHI
4859
```

Fig. 4.2 The Sequence of the 186 DNA from 79.6 to 95.7% of the chromosome (the BGLBAM fragment)

Fig. 4.3 Codon preference plot of the *l*-strand of the BGLBAM fragment

The codon preference plot was constructed by the method given by Gribskov *et. al.* (1982), using the *E. coli* codon frequency table given by Chen *et. al.* (1982).

A, B and C represent the plots for frames 1, 2 and 3, respectively.

The median line represents the average of the plot values obtained by scanning a computer generated random sequence which had the same base composition as that of the BGLBAM fragment. The line above this equals the average value plus standard deviation (SD). Asteriks mark the positions of nonsense codons in each frame, and are marked along the line representing the average value minus SD.

D. The relative organizations of ORFs.

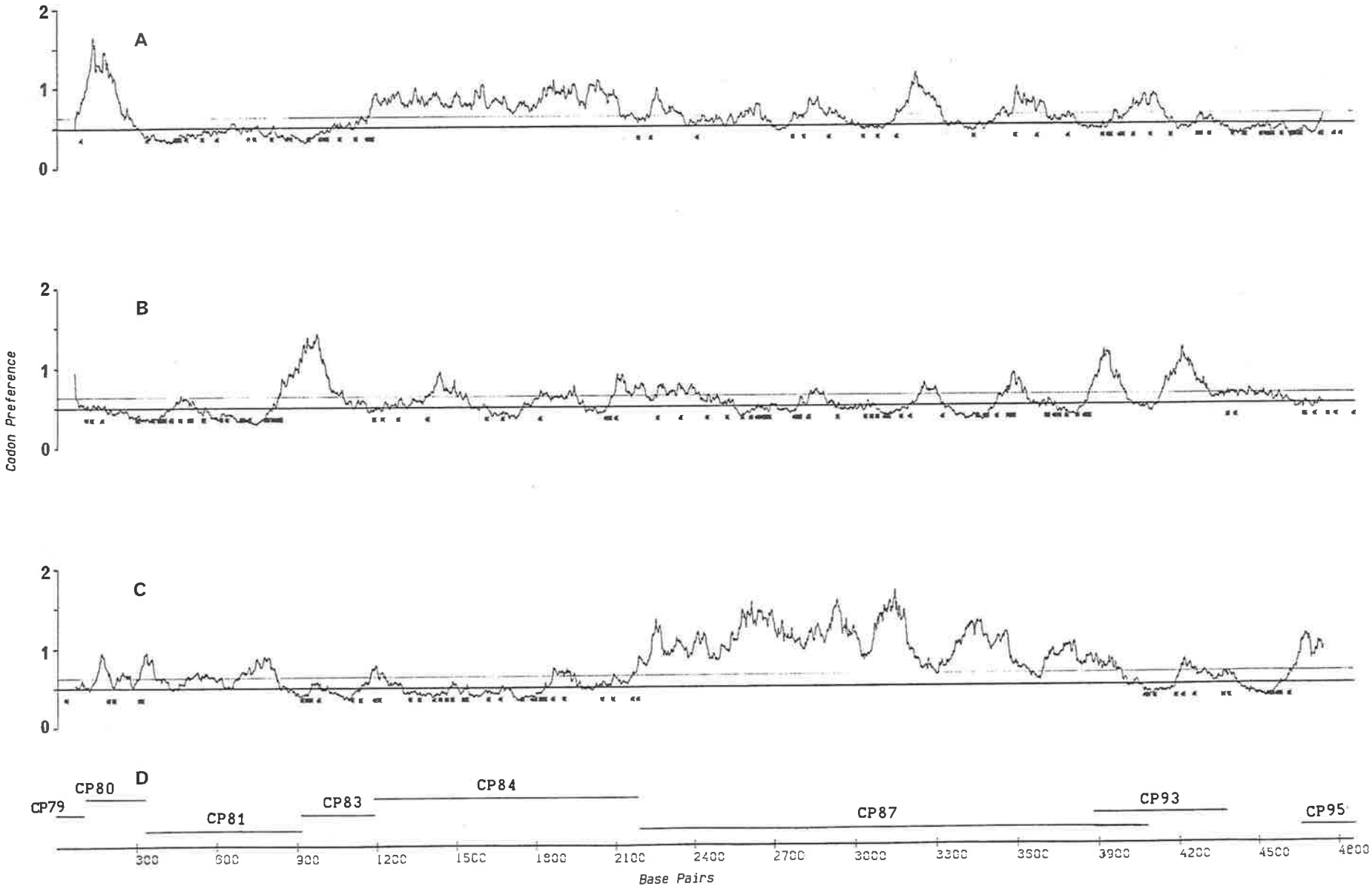
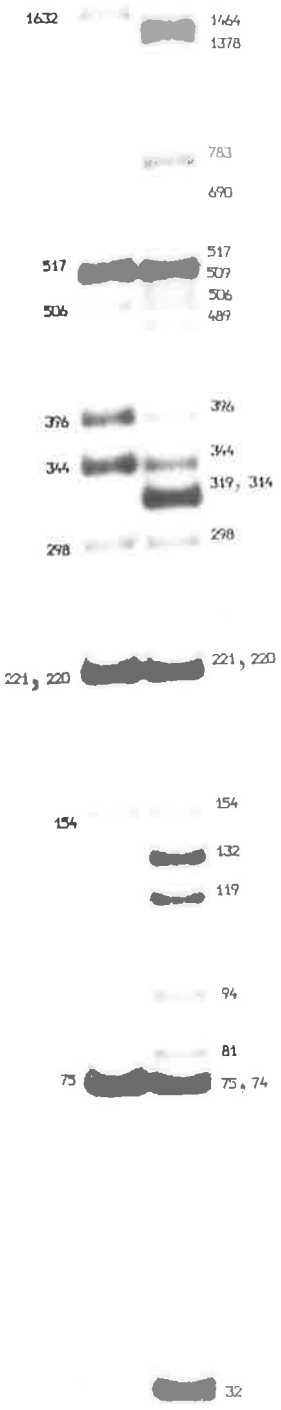

A
B
C
D
2
1
0
Codon Preference
CP79
CP80
CP81
CP83
CP84
CP87
CP93
CP95
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
Base Pairs

Fig. 4.4 Restriction fragments generated by digesting pEC701 and pBR322 with *HinfI*

Plasmid DNA was digested with *HinfI* and end-labelled with ^{32}P *dCTP*. The fragments were separated on a 5% polyacrylamide gel (Acrylamide:Bis ratio, 50:1) and autoradiographed at room temperature.

Sizes of fragments (bp) and given alongside the tracks, were determined from the sequences of pBR322 (Sutcliffe, 1979) and of pEC701 (deduced from the sequences of BGLBAM fragment and pBR322).

pBR pEC
322 701



4.3 ANALYSIS OF THE SEQUENCE

4.3.1 Identification of the Potential Genes

Essentially three criteria were employed to identify the potential genes on the sequence. Firstly, open reading frames (ORFs) of at least 50 amino acids preceded by initiator codons were searched for. Presence of suitable ribosome binding sites were then looked for before the initiator codons of these ORFs. Finally these reading frames were compared with the *E.coli* codon usage Table (Gribskov *et.al.* 1984; Chen *et.al.* 1982) to assess their coding potential.

A summary of the potential genes so identified on the BGLBAM fragment is given in Table 4.1 and Fig. 4.5. All these genes were found on the *l*-strand, and the *r*-strand did not contain ORFs which could be considered as potential genes. All genes had above average preference for the commonly used *E.coli* codons as indicated by the plot in Fig. 4.3. These potential genes were assigned the names *CP79* to *CP95* (CP = Computer Protein) based on the percentage of the 186 chromosome where they initiated. The initiation codon ATG was seen for all genes except for *CP93* which had GTG as the start codon. Ribosome binding sites for the potential genes were identified by employing one or more of the rules proposed by Stormo *et.al.* (1982; 1982a) and it was found that all genes except CP87 had suitable ribosome binding sites. (Table 4.1)

Since the genes *CP80* and *CP93* were completely contained within the BGLBAM fragment they must be considered as potential replication genes, whereas *CP79* and *CP95* were only partially contained on this fragment and therefore represented some functions not associated with replication.

Table 4.1 – Summary of the potential genes on the BGLBAM sequence

ORF	Start*	End#	No of aa	MW	RBS**
CP79*	-125	107	77	8817	a,b,c
CP80	109	334	76	8514	c
CP81	333	915	194	21467	c
CP83	914	1187	92	10408	a,b
CP84	1186	2188	334	37805	a,b
CP87	2187	4068	627	72019	-
CP93	3884	4376	165	18287	a
CP95*	4656		68+		a,b,c

The first base in the start and stop codons are given

* The ORFs CP79 and CP95 are not fully contained on the BGLBAM sequence. The start position of CP79 could, however, be obtained from the sequence to the left of *BglII* (H. Richardson, personal communication) whereas no sequence information is available to the right of *BamHI* and therefore the end position of CP95 is unknown at present.

** Computer search of the sequence for RBS gave values above the threshold limit for the following matrices given by Stormo *et. al.* (1982b).

a. W71

b. W101

c. Rule 6

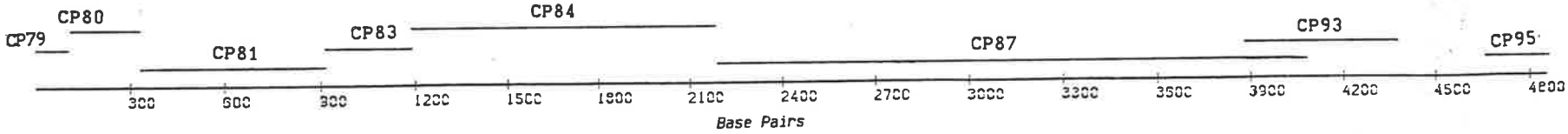
None of the three rules could predict the RBS of CP87.

Fig. 4.5 Potential genes encoded by the BGLBAM fragment

The solid line represents the sequence of the BGLBAM fragment (4859 bp) and the horizontal bars above this line represents the potential genes.

Lengths of the horizontal bars correspond to the length of the potential genes, starting from the first base in the initiator codon and ending on the first base in the terminator codon.

The potential gene *CP79* starts from an ATG codon located at -125 bases from the *BglII* site (H. Richardson, personal communication).


CP79
CP80
CP81
CP83
CP84
CP87
CP93
CP95
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
Base Pairs

4.3.2 Identification of geneA

Although genetic mapping of amber alleles indicated that *geneA* spanned the two *PstI* sites located at positions 1490 and 2362 of the sequence (84.6%* and 87.5%*, respectively, of the 186 chromosome; Finnegan and Egan, 1979), no single ORF was seen spanning these two sites. As two alleles of *geneA* (*Aam24* and *Aam33*) had been mapped between the *PstI* and *EcoRI* sites (2362 and 3732; Finnegan and Egan, 1979) it appeared that they occurred in the reading frame of *CP87*. Five alleles, mapped in between the two *PstI* sites (1490 and 2362), could occur in either *CP84* or *CP87* whereas the only *Aamber* allele (*Aam43*) mapped to the left of *PstI* at 1490 could be in *CP84* or any other gene situated to the left, but not in *CP87*. This raised the possibility that *geneA* was actually represented by more than one cistron.

4.4 *GeneA* IS *LA* PLUS *RA*

4.4.1 Mapping of *Aamber24*

As the *Aam24* allele had been mapped in between *PstI* and *EcoRI* (2362 and 3732), to locate this mutation the entire section between these two sites was sequenced by using specific primers after constructing an M13 clone of the *BglII-BamHI* fragment taken from 186*cItspAam24* phage DNA. The mutation was found to be an AAG to TAG transversion at position

---

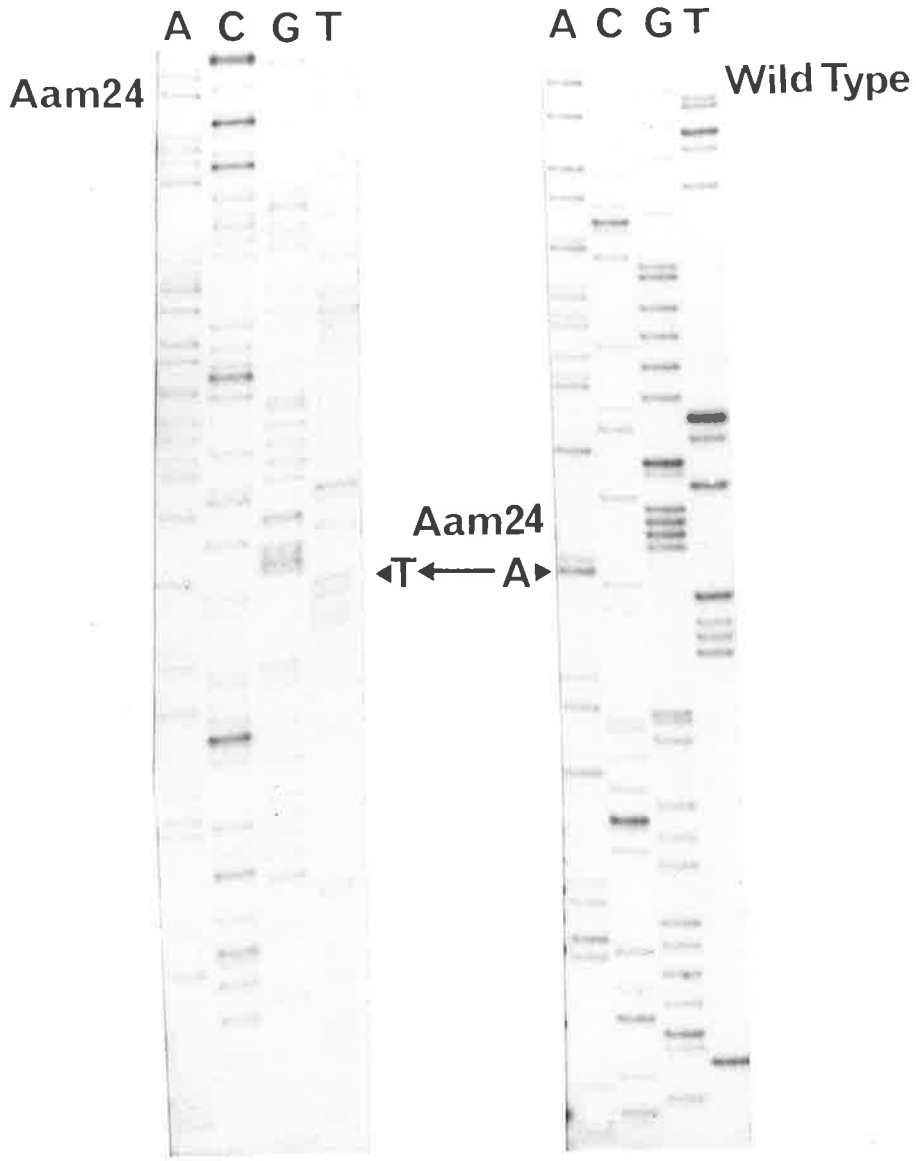
* Locations of restriction sites have been re-defined after obtaining the DNA sequence, and therefore the new values will be used here-after. The PstI sites at 84.6% and 87.5% refer to the sites at 83.8 and 87.0% as described in Finnegan and Egan (1979).

Fig. 4.6 The DNA sequence showing the *Aam24* mutation

The wild type sequence was obtained from a *HpaII* clone (2802 to 3281) by sequencing with the universal primer. The *Aam24* mutation was located by sequencing a *BglII* to *BamHI* clone with the *HhaI* fragment (2837 to 2870) as the primer. Both sequences are from the *l*-strand and therefore correspond to the sequence of the *mRNA*.

The change, A to T transversion, was found at 2985, and was in the reading frame of *CP87* (*RA*).

An example of the kind of 'compressions' observed during the sequencing is seen two bases below the triple T residues in the wild type. The sequences in the compressed area, as read from the *r*-strand, is 5' CGCGCGAATTT 3'.


A C G T
Aam24
A C G T
Wild Type
Aam24
T ← A

2985 of the sequence and the amber codon was in the reading frame of *CP87* (Figure 4.6). No other base in the coding region of this gene was found to be different from the wild-type sequence. Since an *Aamber* mutation was located in the reading frame of *CP87* this gene must be one of the cistrons of *geneA*, and therefore, to retain the original nomenclature of genes on 186, it was renamed as *RA* (for *R*ight *A*).

4.4.2 Mapping of Aam11

Genetic mapping of the *Aam11* allele showed that this mutation was situated in between the two *PstI* sites (positions 1490 and 2362) on the BGLBAM fragment (Finnegan and Egan, 1979). The position of *Aam11* mutation was therefore found by sequencing this *PstI* fragment taken from 186*citspAam11* phage DNA. The change was found at 1912 and was a GAG to TAG transversion which caused the occurrence of an amber codon in the reading frame of *CP84* (Fig. 4.7). Since the presence of *Aam11* mutation identified CP84 as another cistron of *geneA* this gene was renamed as *LA* (for *L*eft *A*).

4.4.3 Mapping of Aamber43

The *Aam43* allele has been genetically mapped to the left of *PstI* site at 1490 (84.6% of the chromosome; Finnegan and Egan, 1979). It was therefore located by sequencing an M13 clone carrying the *PstI* fragment (77.6 to 84.6% of 186 chromosome), taken from 186*cItspAam43* DNA, that gave sequence reading towards the *BglII* site. The change was an AAG to TAG transversion at 1246 which resulted in an amber codon in the reading frame of *LA* (Fig. 4.8). The *Aam43* mutation therefore again confirmed that *LA* was another cistron of *geneA*.

Fig. 4.7 The DNA sequence showing the *Aam11* mutation

The wild type sequence was obtained from a *HpaII* clone (1754 to 1930) by sequencing with the universal primer. The *Aam11* mutation was located by sequencing a *PstI* clone (1490 to 2362) with the universal primer.

Both sequences are from the *l*-strand and therefore correspond to the sequence of the mRNA.

The change, G to T transversion, was found at 1912, and was in the reading frame of *CP84* (*LA*).

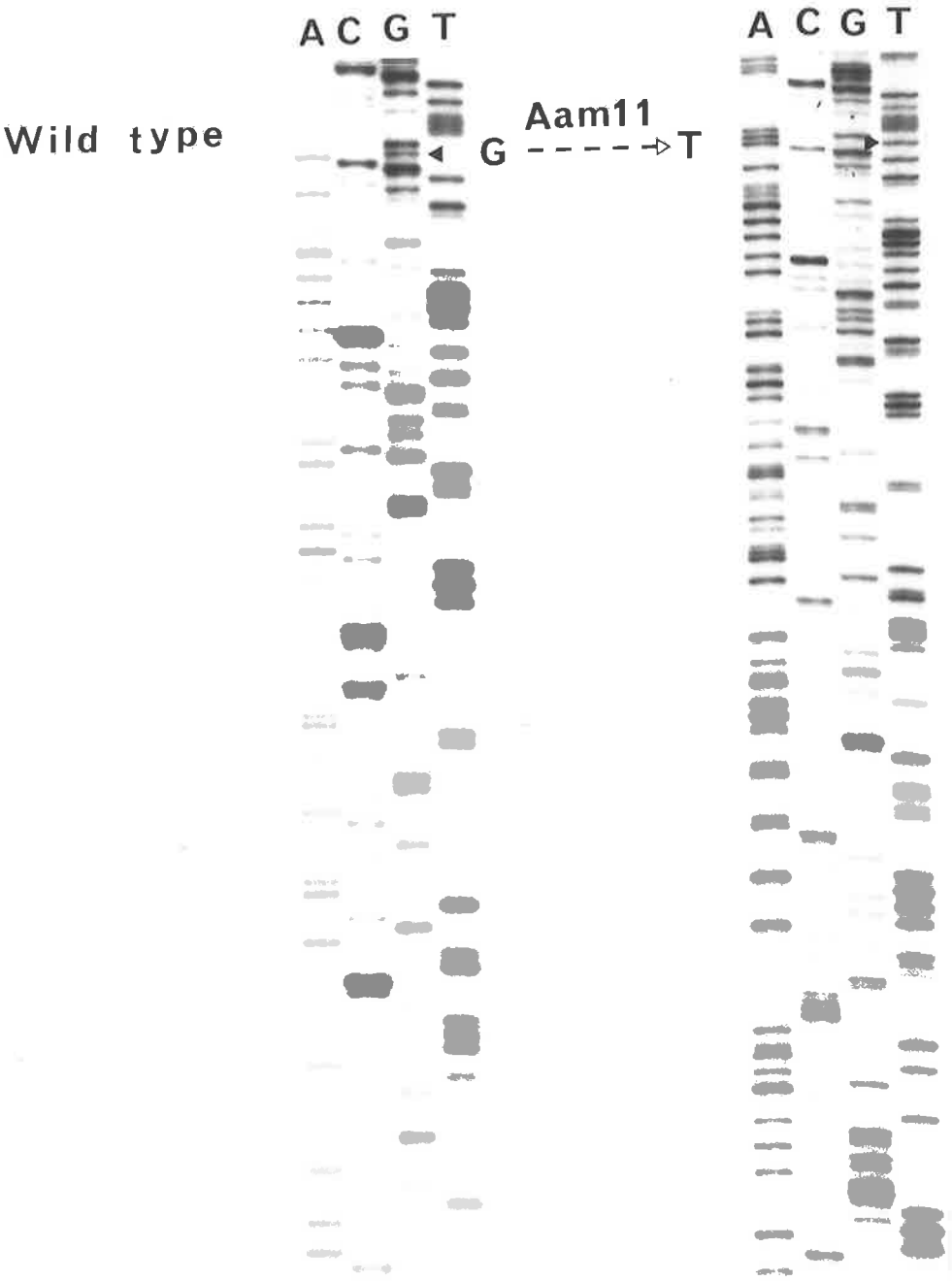
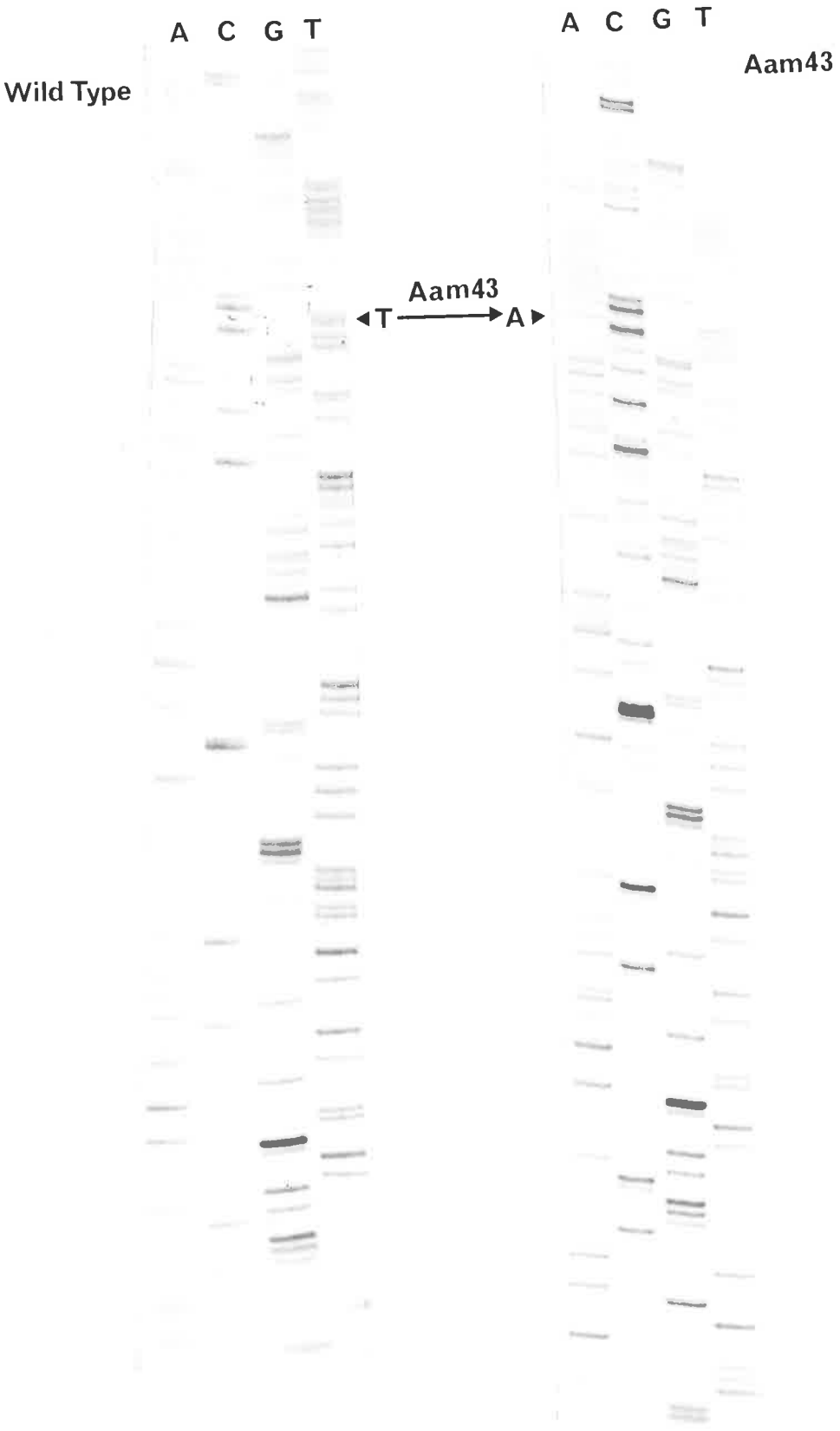

A C G T
Wild type
Aam11
G ----> T
A C G T

Fig. 4.8 The DNA sequence showing the *Aam43* mutation

The wild type sequence was obtained from a *PstI* fragment (77.3 to 83.7% of 186 chromosome) by sequencing with the universal primer. The *Aam43* mutation was located by sequencing the same *PstI* fragment, taken from 186*cItspAam43* phage, with the universal primer.

Both sequences are from the *r*-strand and therefore are complementary to the sequence of the mRNA.

The change, A to T transversion (T to A in figure), was found at 1246, and was in the reading frame of *CP84* (*LA*).


A C G T
A C G T
Aam43
Wild Type
Aam43
T
A

4.5 IDENTIFICATION OF THE GENE PRODUCTS

In order to determine whether the potential genes identified on the sequence were actually expressed *in vivo* the pattern of protein bands produced by pEC701 was investigated in collaboration with A. Puspurs. Figure 4.9 gives the protein bands which could be identified by labelling minicells habouring pEC701. The estimated molecular weights of these protein bands agreed well with the predicted molecular weights (Table 4.2) and so the reality of the genes was indicated. No protein band corresponding to the size predicted for CP93 (18 kd) could however be found. In contrast to what was expected for a large protein, the band of *RA* was very faint. This did not appear to be due to a failure of the protein to enter the gel as no significant amount of radioactivity was visible at the top of the gel. Low production or high rate of degradation *in vivo* of the RA protein could explain the faint band for this protein. A similar reason for the absence of the 18.3 kd potential protein product of the distal gene *CP93* seems likely.

4.5.1 Identification of the LA and RA products

In order to identify the protein bands assigned to *RA* and *LA*, minicells, carrying pEC703 (= pEC701 *Aam24*) or pEC450 (= pEC701 *Aam11*), were labelled (Fig. 4.10). The *Aam11* mutation in pEC450 was found to cause the disappearance of the 38 kd band originally assigned to LA, thereby confirming the identity of this protein band. The 7.5 kd amber fragment predicted from the sequence was found to appear when the 38 kd band disappeared (shown by arrow).

Table 4.2 - Molecular weights of the potential genes

ORFs	Calculated from computer transln.	Estimated from Protein gels
CP80	8.514 Kd	6.0-8.0 kd
CP81	21.467 "	22.5 "
CP83	10.509 "	12.3 "
CP84	37.805 "	38.0 "
CP87	72.019 "	69.0 "
CP93	18.287 "	-- "
CP95*	15.797 "	16.5 "

* This is a fusion protein containing pBR sequences

Fig. 4.9 The translation products directed by pEC701 in minicells

Minicells, prepared from DS410(pEC701) or DS410(pEC702) according to the method of Reeve (1979), were labelled with L-[^{35}S]methionine as described by Giphart-Gassler *et. al.* (1981). Electrophresis through 14-20% gradient ployacrylamide-SDS gel and fluorography were as described by Reeve and Shaw (1979). The gel was autoradiographed at 80°C. Tracks:

1. Translation products in minicells carrying pEC701
2. Translation products in minicells carrying pBR322
M. Markers (all methylated with [methyl-^{14}C]; purchased from New England Nuclear, Boston, Mass., U.S.A.):

		MW
a. Bovine serum albumin	-	69000
b. Ovalbumin	-	46000
c. Carbonic Anhydrase	-	30000
d. Lactoglobulin, A	-	18367
e. Cytochrome C	-	12300
f. Insulin	-	5766

(Note: This photograph was prepared, with consent, from the results of work done by A. Puspurs.)

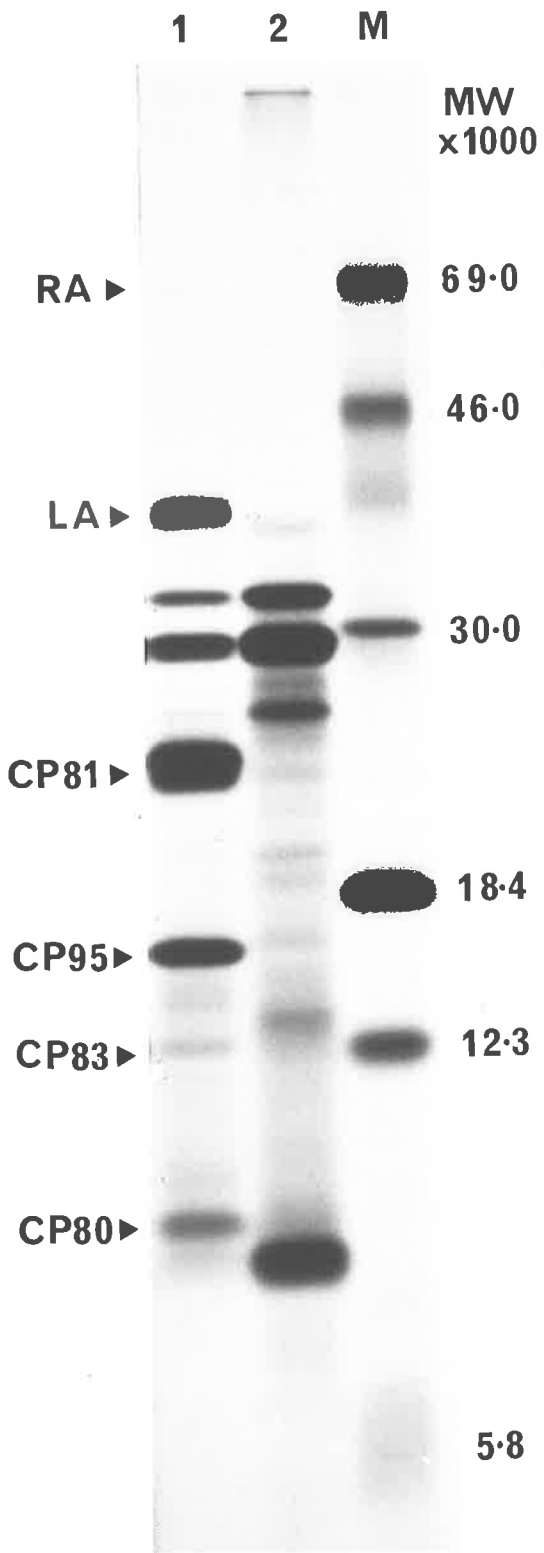

1
2
M
MW
x1000
RA ▶
69·0
46·0
LA ▶
30·0
CP81 ▶
18·4
CP95 ▶
CP83 ▶
12·3
CP80 ▶
5·8

Fig. 4.10 The translation products directed by pEC701, pEC703, pEC450 and pBR322 in minicells.

Derivatives of DS410 harbouring pEC701, pEC703 (= pEC701 *Aam24*), pEC450 (= pEC701 *Aam11*) or pBR322 were used for labelling the translation products.

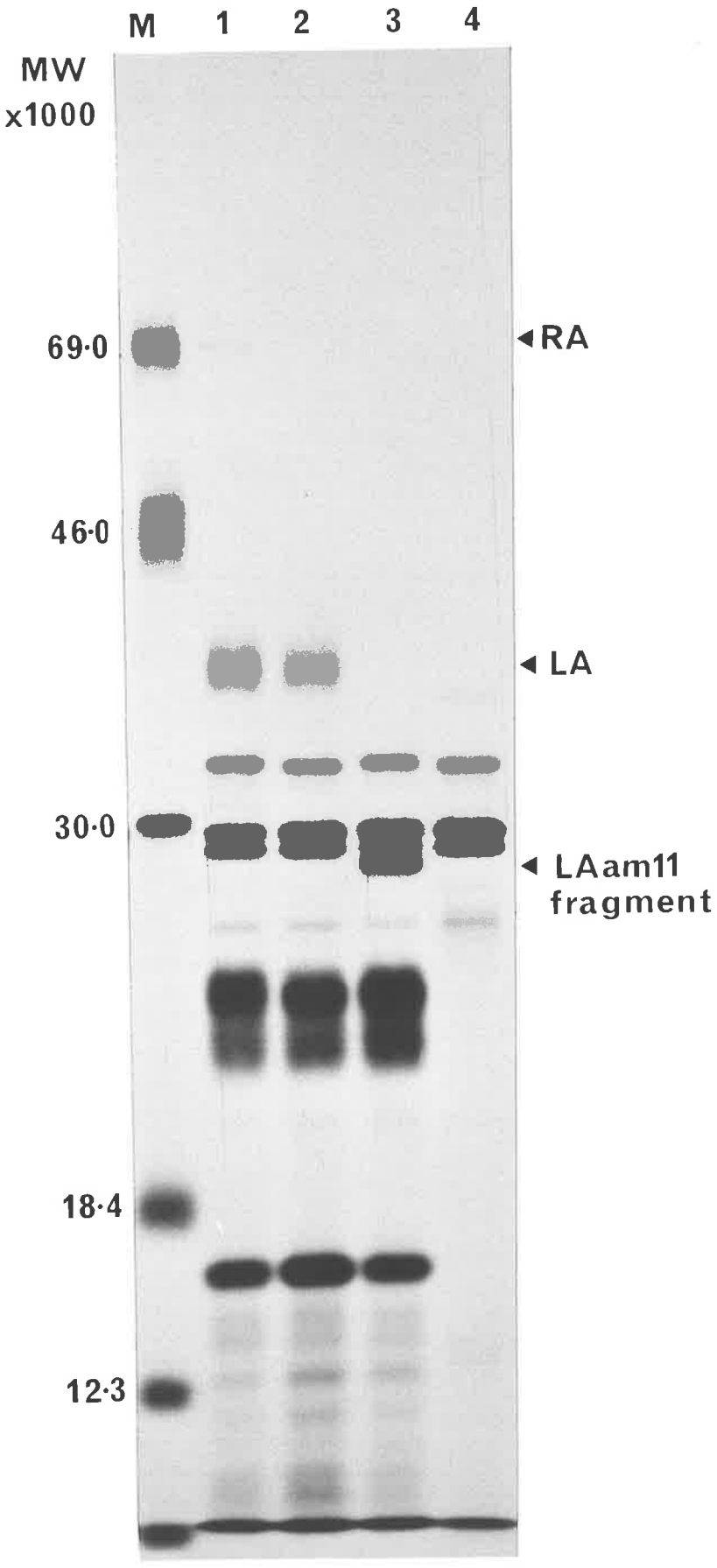
Preparation of minicells and labelling were done exactly as given in the legend for Fig. 4.9. Electrophoresis was done through 12.5% polyacrylamide-SDS gel and autoradiographed after fluorography.

Tracks:

M. Markers (see legend in Fig. 4.9)

1. Minicells carrying pEC701
2. Minicells carrying pEC703
3. Minicells carrying pEC450
4. Minicells carrying pBR322

(Note: This photograph was prepared, with consent, from the result of work done by A. Puspurs; Thanks to S. Williams for providing pEC450.)


M
1
2
3
4
MW
x1000
69·0
46·0
30·0
18·4
12·3
RA
LA
LAam11
fragment

The *Aam24* mutation similarly caused the disappearance of the 69 kd band assigned to *RA*. Although a 31 kd amber fragment had been predicted from the sequence, no such band could be detected in this case. The failure to detect the amber fragment could possibly be due to the very faint nature of the *RA* band or that the amber fragment was masked by the pBR bands which occurred in the region of 31 kd.

An interesting feature of *Aam11* mutation was that in addition to causing the disappearance of *LA* protein band this mutation caused the disappearance of the band corresponding to *RA*. It therefore appeared that this mutation had a polar effect on the synthesis of *RA*. This polar effect could stem either from a direct involvement of *LA* protein in the synthesis of *RA* or from transcriptional/translational polarity of the amber mutation. Possiblility of a double mutation was unlikely as the reversion rate of the *Aam11* mutation was close to that for a single base change ($2x10^{-6}$). Sequencing the entire length of *RA* was therefore not performed to locate another possible mutation in its reading frame.

4.5.2 Identification of the products of other genes

Since no other amber mutation has been genetically mapped to the left of *PstI* at 1490 or to the right of *EcoRI* at 3732, positive identification of the protein products of the other genes carried on BGLBAM fragment was not possible. The bands assigned to genes *CP80*, *CP81* and *CP83* were however absent when pEC705, a plasmid that lacked DNA sequences to the left of position 1666, was used to label the proteins (Fig. 4.11). This indirectly suggested that the bands assigned to these

Fig. 4.11 The translation products directed by pEC701, pEC705 and pBR322 in minicells

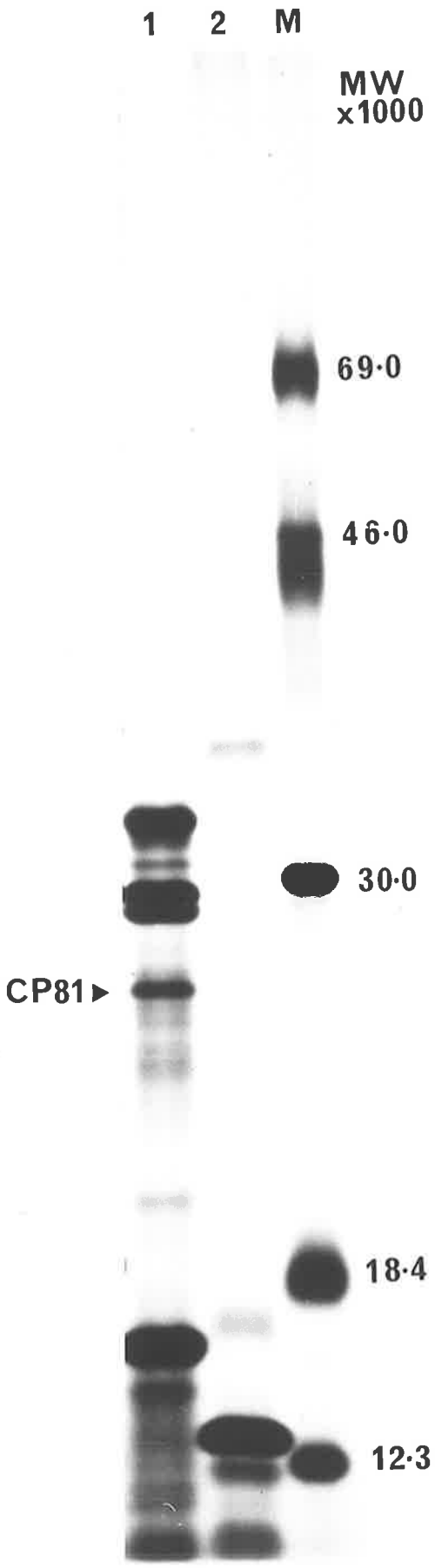
Derivatives of DS410 harbouring pEC701, pEC705 or pBR322 were used for labelling the translation products.

Preparation of minicells and labelling were done exactly as given in the legend for Fig. 4.9. Electrophoresis was done through 12.5% polyacrylamide-SDS gel and autoradiographed after fluorography.

Tracks:

1. pEC701
2. pEC705
M. Markers (see Fig. 4.9 for description)

(Note: This photograph was prepared, with consent, from the result of work done by A. Puspurs.)


1
2
M
MW
x1000
69·0
46·0
30·0
CP81
18·4
12·3

genes actually belonged to them. Identity of *CP93* still remained doubtful.

4.6 OTHER FEATURES OF THE SEQUENCE

4.6.1 Restriction sites

A fine restriction map of the BGLBAM fragment showing known restriction sites as well as sites of other enzymes which cut this fragment once is given in Fig. 4.12. Table 4.3 gives the positions of all the commonly used restriction enzymes. Several sites unique on the BGLBAM fragment were also unique on the whole pEC701 DNA. Sites of *PstI*, *EcoRI* and *SacI*, which have previously been mapped on the whole phage DNA (Saint and Egan, 1979), appeared on the sequence at the expected locations and were separated from each other by the expected number of bases.

Biased distributions of the sites of *Sau3AI* ("GATC"), which appeared only twice as against an expected number of 16, and of *BssHII* ("GCGCGC"), which occured seven times instead of one, were seen on the sequence. No other tetra- or penta-nucleotide sequences was seen at significantly higher or lower frequencies.

The site for *Not1* ("GCGGCCGC") was found once on the sequence and was later found to be unique on the phage chromosome.

4.6.2 Origin of replication

The sequence CACTAT, which is a six-base consensus sequence found in or near the origin of replication of phages

Table 4.3 Locations of Restriction Sites

Enzyme	Recognition Sequence	No.of sites	Locations
AccI	GTCTAC	1	3897
AccI	GTATAC	1	3675
AcyI	GGCGTC	1	2543
* AhaII	GGCGTC	1	2543
AhaIII	TTTAAA	2	710 774
AluI	AGCT	21	54 99 299 596 1012 1062 1812 1904 2076 2241 2415 2527 3079 3121 3237 3274 3685 4324 4377 4382 4530
AvaI	CCCGAG	1	1575
AvaI	CTCGGG	1	2471
AvaII	GGACC	1	3798
* BamHI	GGATCC	1	4854
* BanII	GAGCTC	1	3120
BbvI	GCAGC	3	179 909 4322
BbvI	GCTGC	14	272 810 1199 1489 1714 1871 1972 2134 2297 2361 2862 3275 3753 4319
* BglI	AGATCT	1	1
Bsp1286	GGGCAC	1	3256
Bsp1286	GAGCAC	2	1578 4538
Bsp1286	GAGCTC	1	3120
Bsp1286	GTGCTC	2	1691 2468
BssHII	GCGCGC	7	175 241 1942 2597 2711 3459 3624
BstNI	CCAGG	2	765 3715
DdeI	CT-AG	14	294 1115 1620 1694 1809 1901 2117 3435 3472 3723 4079 4115 4379 4584
* EcoK	GCAC------GTT	1	791
EcoPI	AGACC	5	150 999 3422 4605 4754
EcoPI	GGTCT	1	1073
* EcoRI	GAATTC	1	3732
EcoRII	CCAGG	2	765 3715
* EcoRV	GATATC	1	2625
FnuDII	CGCG	36	65 139 154 174 176 242 756 1292 1711 1943 1945 2024 2105 2204 2274 2288 2596 2598 2621 2643 2712 2838 2871 2924 2952 2971 2973 3267 3309 3460 3625 4292 4346 4572 4834 4836

Table 4. 3 Continued

Enzyme	Recognition Sequence	No.of sites	Locations
Fnu4HI	GC-GC	35	47 63 179 245 272 757 807 810 909 1199 1489 1714 1871 1972 2025 2134 2272 2289 2297 2361 2496 2619 2839 2842 2859 2862 2965 3070 3275 3338 3621 3738 3753 4319 4322
FokI	GGATG	3	868 1737 2789
FokI	CATCC	5	973 3398 3635 4150 4413
HaeI	AGGCCT	1	3248
HaeI	AGGCCA	1	3475
HaeI	TGGCCT	1	1444
HaeI	GGCGCT	1	3719
HaeI	AGCGCT	2	462 1647
HaeI	AGCGCC	1	2962
HaeIII	GGCC	17	7 49 111 527 725 1372 1445 1928 2618 2654 2841 3086 3249 3476 3982 4255 4767
HgaI	GACGC	5	115 888 2851 3766 4344
HgaI	GCGTC	4	2544 2938 3310 4573
HgiAI	GTGCTC	2	1691 2468
HgiAI	GAGCAC	2	1578 4538
HgiAI	GAGCTC	1	3120
HhaI	GCGC	38	61 138 175 177 241 243 463 564 589 730 1293 1648 1712 1942 1944 2339 2597 2599 2644 2711 2713 2813 2837 2870 2951 2963 2972 3268 3459 3461 3624 3626 3663 3720 4004 4291 4571 4835
HincI	GTCAAC	2	2389 4829
HincI	GTTGAC	2	3986 4102
HincI	GTTAAC	1	3941
HinfI	GAATC	4	89 1360 1499 4622
HinfI	GACTC	6	872 1367 1618 2013 3391 4113
HinfI	GAGTC	2	1699 4081
HinfI	GATTC	2	1192 1266
* HpaI	GTTAAC	1	3941
HpaII	CCGG	23	235 259 866 1002 1316 1375 1532 1754 1930 2174 2586 2671 2716 2802 3281 3369 3486 3666 3938 4215 4231 4257 4773
HphI	GGTGA	7	423 935 3558 3579 3668 4307 4701
HphI	TCACC	4	951 1084 1702 2941
MboI	GATC	4	2 2994 4049 4855

Table 4.3 Continued

Enzyme	Recognition Sequence	No.of sites	Locations
MboII	GAAGA	15	148 381 1051 1160 1462 1795 2062 3000 3006 3285 3372 3420 3873 3953 4120
MboII	TCTTC	1	4456
MluI	ACGCGT	3	2104 2203 3308
MnlI	CCTC	4	2582 3846 3907 4008
MnlI	GAGG	17	193 214 323 519 1019 1130 1732 1735 1933 2787 2826 3417 3474 4512 4586 4645 4719
* MstI	TGCGCA	1	2812
NciI	CCGGG	4	259 1316 3281 4231
NciI	CCCGG	6	234 1374 1753 2585 3485 3937
NciI	CCGGG	4	259 1316 3281 4231
NciI	CCCGG	6	234 1374 1753 2585 3485 3937
* NdeI	CATATG	1	3348
* NotI	GCGGCCGC	1	2839
PstI	CTGCAG	3	1490 2362 4320
RsaI	GTAC	12	1079 1288 1624 1639 2121 2696 2705 3828 3832 4032 4425 4589
* SacI	GAGCTC	1	3120
SacII	CCGCGG	2	2023 2620
Sau3AI	GATC	4	2 2994 4049 4855
Sau96I	GG-CC	3	1372 3086 3798
ScrFI	CC-GG	12	234 259 765 1316 1374 1753 2585 3281 3485 3715 3937 4231
SfnAI	GATGC	8	223 429 737 804 1738 3750 3903 4782
SfnAI	GCATC	6	2267 2476 3019 3323 3886 4218
SphI	GCATGC	2	1874 2300
* SstI	GAGCTC	1	3120
SstII	CCGCGG	2	2023 2620
* StuI	AGGCCT	1	3248
TaqI	TCGA	17	886 927 1029 1383 1547 2011 2506 2998 3379 3538 3553 3848 3909 4062 4194 4282 4560
XhoII	GGATCC	1	4854
XhoII	GGATCT	1	2993
XhoII	AGATCT	1	1
XmaIII	CGGCCG	2	48 2840
* XmnI	GAA----TTC	1	1666

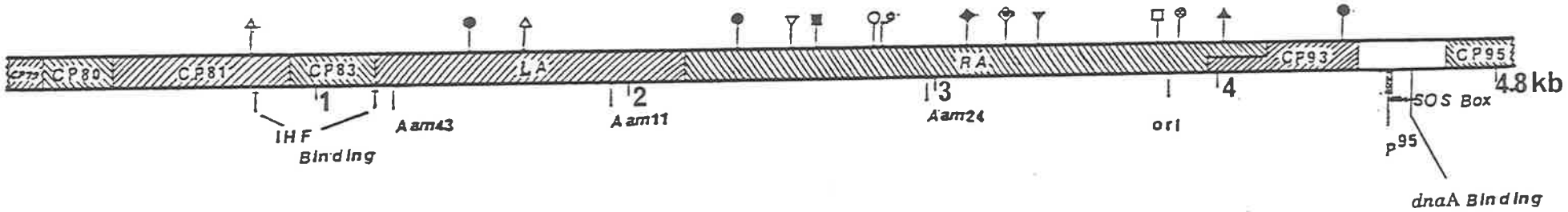
* - Enzymes which cut BGLBAM fragment only once each

Fig. 4.12 Summary of the sites on the BGLBAM fragment

Sites of the restriction enzymes which cleave the 79.6 to 96% of 186 DNA (BGLBAM fragment) only once are shown, together with the sites of the three *PstI* sites.

The hatched areas represent the coding regions of the potential genes.

The positions of the amber codons and of the other signal sequences are also shown.



Ava II Aha II EcoK EcoRI EcoRV Hpa I Mst I

Nde I Not I Pst I Sac I Stu I Xmn I

(Kornberg, 1982), was found at position 3759 (92.0%). The EM *ori* (Chattoraj and Inman, 1973) has been proposed to lie between 3190 and 4270 of the BGLBAM fragment (92.9 ± 1.8% of 186 chromosome) and therefore this site is within the limits of *ori*. The sequence CACTAT will therefore be considered as the putative *ori* of 186 until the exact origin is located by other means.

4.6.3 Promoters and Terminators

Consensus sequences for the -35 and -10 regions of prokaryotic promoter were found at positions 4458 and 4482, respectively, (Figure 4. 2) and this promoter, termed p95, was found active *in vitro* by M. Pritchard (personnel communication). A stem-loop structure (ΔG = -4.12 Kcal/mol) followed by a run of 7 Ts was seen immediately to the left of this promoter and was thought to represent a transcription terminator (Fig. 4.13). Another very stable stem loop structure (ΔG = -12.79 Kcal/mol) followed by a run of 3 Ts was located at 2271 of the sequence and so was within the reading frame of RA (Figure 4.14). No promoter was however seen in the area following this possible terminator.

4.6.4 Protein binding sequences

A very significant homology (18/20) with the 20-base consensus sequence for *lexA* binding site ("SOS Box"; Little, *et.al.*, 1981; Sancar *et.al.*, 1982; Sancar *et.al.*, 1980; Walker, 1984) was found at 4482. This "SOS Box" overlapped the -10 region and the mRNA initiation site of p95. (Fig. 4.2).

Fig. 4.13 Potential Secondary structure before p95

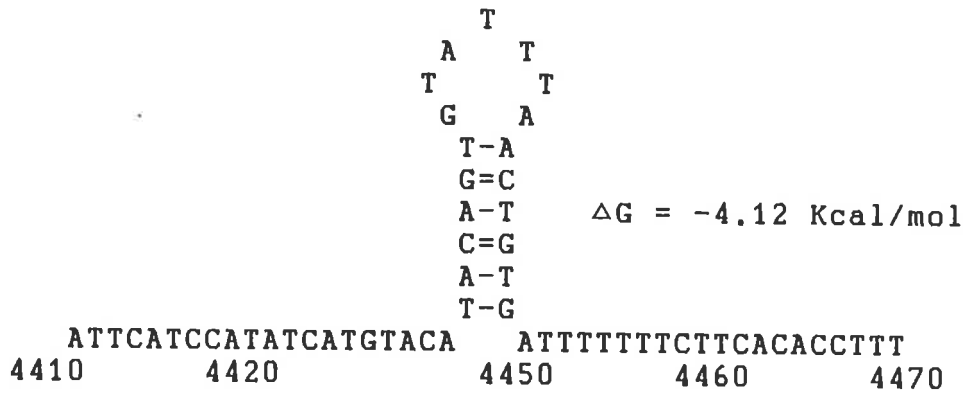
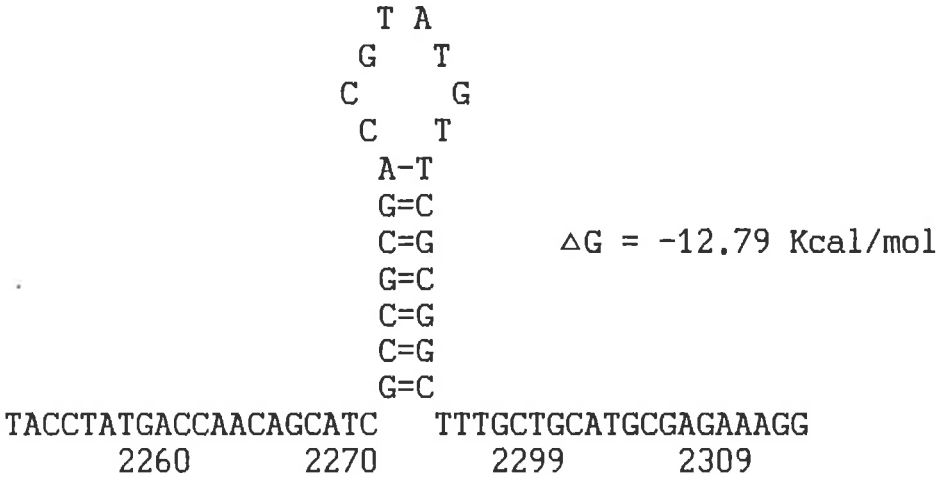

T
A T
T T
G A
T-A
G=C
A-T
C=G
A-T
T-G
ΔG = -4.12 Kcal/mol
ATTCATCCATATCATGTACA ATTTTTTTCTTCACACCTTT
4410 4420 4450 4460 4470

Fig. 4.14 Potential Secondary structure in the coding region of *RA*



$\triangle G = -12.79$ Kcal/mol

The consensus sequence for the binding of Integration Host factor (IHF; 5'-AANNNPuTTGAT-3'; Craig and Nash, unpublished) was seen at 796 and 1184. The site found at 1184 overlapped the translation initiation site of *LA*. (Fig 4.2).

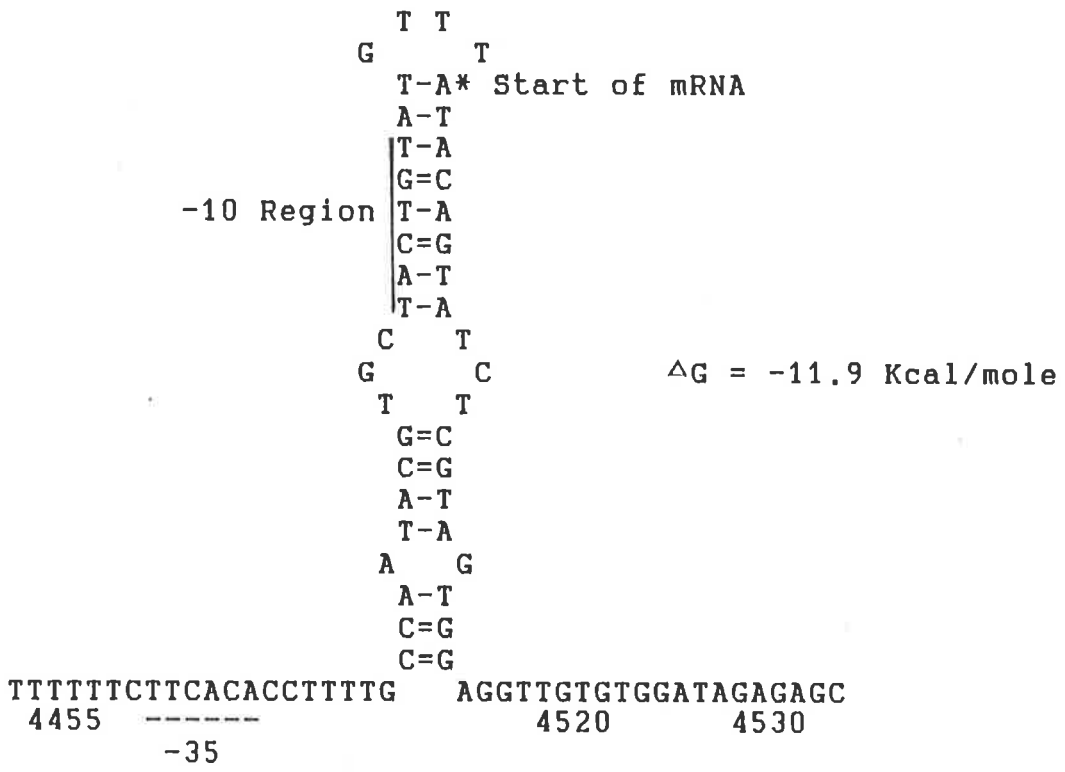
4.6.5 Secondary Structures

No extensive direct or indirect repeats were found on BGLBAM sequence and the sequence near the *ori* region was relatively free of secondary structures. A stable stem loop structure (ΔG = -11.9 Kcal/mol) was found at 4470 with its stem spanning the -10 region of p95. (Fig. 4.15).

4.7 DISCUSSION

The DNA sequence of the BGLBAM fragment revealed six potential genes and two of these were renamed as *LA* and *RA* after sequence mapping the amber alleles *Aam11* and *Aam43* in the reading frame of *CP84*, and *Aam24* in the reading frame of CP87. Since the protein bands identified as those belonging to *LA* and *RA* agreed well with the molecular weights of proteins predicted from the sequence, the two cistrons for *geneA*, which was genetically defined as a single gene, did not arise as a result of a sequencing error. Possibility of a post-translational modification of protein to give two bands for an otherwise single cistron also seemed unlikely because the relative intensities of the two bands suggested an unequal production of these proteins. Multimers of polypeptides would not remain associated under the conditions in which the gels were run and therefore it appeared unlikely that the 69 kd

Fig. 4.15 Potential Secondary structure covering the -10 region of p95.



$\triangle G$ = -11.9 Kcal/mole

band was actually a dimer of the 38 kd band. Furthermore, the *Aam24* mutation in *RA* caused the disappearance of the 69 kd band but did not affect the 38 kd band which proved that the former was not a dimer of the latter.

Polarity of the *LAam11* mutation on the expression of *RA* was indicated by the disappearance of the 69 kd band together with the 38 kd band when the plasmid used for labelling carried this mutation.

Positive identification of the protein products of the other genes carried on BGLBAM fragment could not be done due to the unavailability of amber mutations. However, the existence of *CP80*, *CP81* and *CP83* was indicated by the absence of their putative protein bands when pEC705, which did not carry the DNA upto 1666 of BGLBAM, was labelled. Reality of the potential gene *CP93* could not be established from minicell studies which might mean that this gene either did not exist or was expressed at a very low level.

Sequence mapping of the *Aamber* alleles in their repective reading frames made *LA* and *RA* the most likely candidates for being the replication genes. As no other function has so far been assigned to the other genes present on the BGLBAM fragment they too are potential replication genes.

Identification of the actual replication gene(s) can now be pursued with the help of the DNA sequence of the BGLBAM fragment.

CHAPTER 5

REPLICATION GENE AND ITS CONTROL

5.1 INTRODUCTION

Though genetic evidence suggested that both *LA* and *RA* were needed for replication, the polarity of *LAam11* on the expression of *RA* raised the possibility that the block in replication associated with this mutation might be due to this polarity rather than a direct involvement of *LA* product in replication. This possibility can be resolved by mutating *LA* without affecting the expression of *RA*. Involvement of the other genes, carried on BGLBAM fragment, in replication can be investigated either by site-directed mutagenesis or by specific deletions of these genes from pEC701.

This chapter describes the identification of *RA* as the only gene essential for the replication of 186 *ori* in pEC701, and also addresses the question of *LA* control on the expression of *RA*. Involvement of *LA* in replication of the phage DNA was investigated by the use of a temperature sensitive revertant of the *LAam11* phage.

5.2 METHODS

5.2.1 Construction of *LAts* phage

Overnight culture of E1011 was subcultured into 10 ml of LB and grown at 30°C to an O.D. of 0.4 and, then, heat induced at 45°C for 10 minutes. The culture was then incubated with aeration at 30°C for 2 hours. From the supernatent of the culture 0.1 ml was plated on E536 and incubated overnight at 30°C. The resulting plaques were screened for temperature sensitivity by tooth picking onto lawns of E536 and E251, and

incubating at 40°C. (The strain E251 is isogenic to E536 except for the *strA* allele and, since previous experience (Hocking, 1977) had shown that the pulse label incorporation was better when E251 derivatives were used, it was decided to test the temperature sensitive *LA* mutant by using lysogens of E251 carrying the *LAts* prophage.) One temperature sensitive plaque was plated at 30°C on E251 and cells from the middle of the plaques streaked at 30°C for isolating lysogens of the phages. The lysogens were tested by cross streaking against 186*Itsp* and 186*vir1* phages.

5.2.2 Pulse labelling

Overnight culture of the strain E4012 (= E251 (186*cItspLAts11*)) was prepared in LB at 30°C. A 1:100 dilution of this culture was made in 5 ml of TPGCAA and incubated with shaking at 30°C to an O.D. of 0.3. Flasks were, then, transferred into a water bath at 40°C and incubation continued with aeration. Samples of 200 ul were withdrawn at intervals of 10 minutes and transferred into Eppendorfs tubes containing 50 ul of the labelling solution (50 uCi/ml ^{3}H-thymidine in TPGCAA) which had been pre-warmed and kept in a 40°C heating block. Pulsing was terminated after 2 minutes by transferring 100 ul of the sample onto GF/A filters and immediately immersing the filters in ice-cold 10% TCA. The filters were washed three times with cold 10% TCA, followed by one washing each in ethanol and ether. They were, then, dried at 65°C for 1 hour and counted in a Packard Scintillation counter after adding scinillation fluid.

5.3 RESULTS

5.3.1 CP93 IS NOT NEEDED FOR 186-SPECIFIC REPLICATION OF pEC701

The reading frame of CP93 starts at 3884 and ends at 4376. Deleting the DNA between *PstI* (position 4320) and *BamHI* (position 4854) from BGLBAM fragment will therefore remove 18 amino acids from the C-terminal end of this gene (Fig. 4.2). It was assumed that this would render CP93 non-functional.

Since there are four *PstI* sites on pEC701, deleting the DNA between 4320 (*PstI*) and 4854 (*BamHI*) depends on the isolation of the right fragment after partial digestion with *PstI*. Alternatively, the same deletion can be accomplished by making use of the unique *HpaI* site in combination with *PstI* and *BamHI* sites. In the latter scheme, the DNA between *HpaI* and *BamHI* is deleted first, followed by the addition of the fragment between *HpaI* and *PstI* (Fig. 5.1).

After digesting pEC701 DNA with *HpaI* and *BamHI*, followed by end-filling and phsophatasing, the larger fragment (Fragment A) was isolated from agarose gel. The *HpaI* to *PstI* DNA fragment was end-filled and isolated from polyacrylamide gel (Fragment B). Fragments A and B were ligated together and introduced into E901. Plasmids from these colonies were tested for the *HinfI* digestion pattern and one was found to give the pattern predicted from the sequence, which confirmed that this plasmid (pEC709) did not contain the DNA between *PstI* and *BamHI* of pEC701 (Fig. 5.2).

The 186-specific replication of this plasmid was indicated by its ability to transform *polA* strain. The high

Fig. 5.1 Diagramatic representation of the construction of pEC709

The pEC701 DNA was digested with *HpaI* and *BamHI* and the larger fragment (fragment A) was isolated from 1% agarose gel after end-filling. A separate lot of pEC701 DNA was digested with *HpaI* and *PstI*, end-filled in the presence of ^{32}P dCTP and electrophoresed on 5% polyacrylamide gel (Ac:Bis ratio of 50:1) to isolate the 379 bp band (fragment B). Fragments A and B were ligated together and used to transform E901.

Thick solid lines represent the fragments which were isolated. Potential genes carried on the 186 DNA fragment of pEC701 are shown inside the circle.

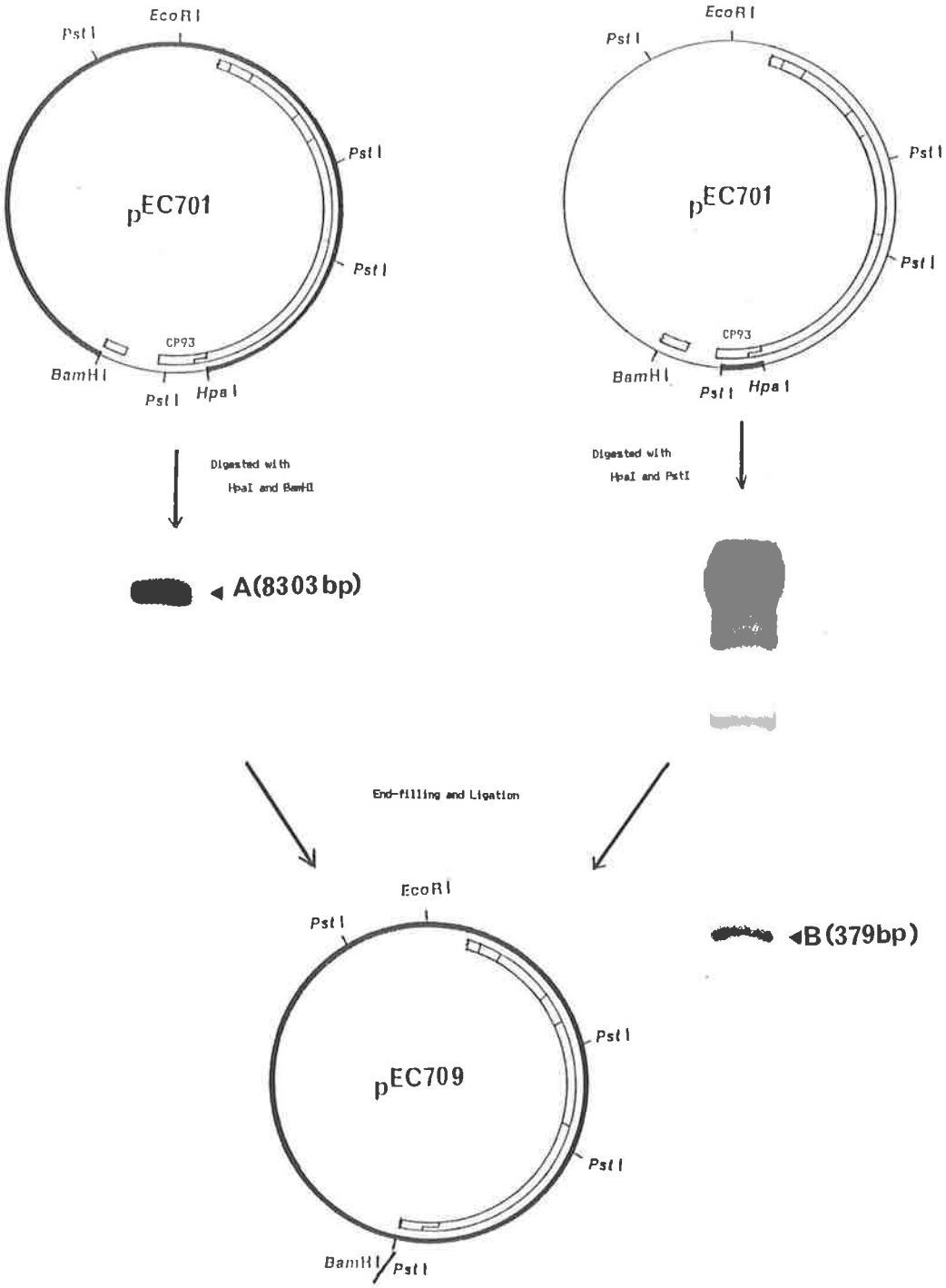

EcoR I
Pst I
Pst I
pEC701
Pst I
CP93
BamH I
Pst I
Hpa I
Digested with
HpaI and BamHI
A(8303 bp)
EcoR I
Pst I
Pst I
pEC701
Pst I
CP93
BamH I
Pst I
Hpa I
Digested with
HpaI and PstI
End-filling and Ligation
EcoR I
Pst I
Pst I
pEC709
Pst I
BamH I
Pst I
B(379bp)

Fig. 5.2 Fragments generated by digesting pEC709 with *HinfI*

HinfI digested pEC701 and pEC709 DNAs were end-filled in the presence of ^{32}P dCTP and separated on 5% polyacrylamide gel (Ac:Bis ratio of 50:1) at a constant 400V.

The solid arrows point to the two fragments (509 and 488 bp, respectively) which must be absent, and the open arrow points to the extra fragment (463 bp) which must appear when the DNA between 4319 and 4859 of the BGLBAM sequence is deleted from pEC701.

pEC701
pEC709

yield of plasmid pEC709 from stationary phase culture was characteristic of pEC701 which was further evidence for its 186-specific replication (Fig. 5.3). However, the colonies carrying this plasmid were only slightly smaller than those produced by pEC702 (Fig. 5.4) and therefore were phenotypically different from those carrying pEC701.

Ability of pEC709 to transform *polA* strains was taken as evidence for its 186-specific replication and therefore it appeared that *CP93* was not required for replication.

5.3.2 *RA* IS DIRECTLY NEEDED FOR 186-SPECIFIC REPLICATION OF pEC701

By virtue of the overlapping of the reading frame of RA with the presumptive *ori*, it appeared to be a very likely candidate for being a replication gene. Furthermore, the *Aam24* mutation located in the reading frame of RA caused the loss of the 186-specific replication of pEC701.

5.3.2.1 Mutating RA at the SacI site prevented replication from phage ori

The unique *SacI* site on pEC701 (position 3120) is situated in the coding region of *RA* and therefore deleting four bases from this site will cause a frameshift mutation in *RA*.

Plasmid DNA was cut with *SacI*, treated with Klenow in the presence of all four dNTPs and the linear form isolated from agarose gel. This was self-ligated and introduced into E536 to give the plasmid pEC707. Plasmid DNA from one of the resulting colonies was used to transform the *polA* strain and found that it lacked the ability to transform these cells.

Fig. 5.3 Comparison of yields of pEC709, pEC702, pEC706 and pBR329

Derivatives of E536 carrying these plasmids were cultured overnight in LB containing the appropriate antibiotics, and plasmid DNAs were prepared from 1 ml of these cultures. This was treated with RNAseA and the entire amount was run on 1% agarose gel.

Tracks:

1. pEC709
2. pEC702
3. pEC706
4. pBR329

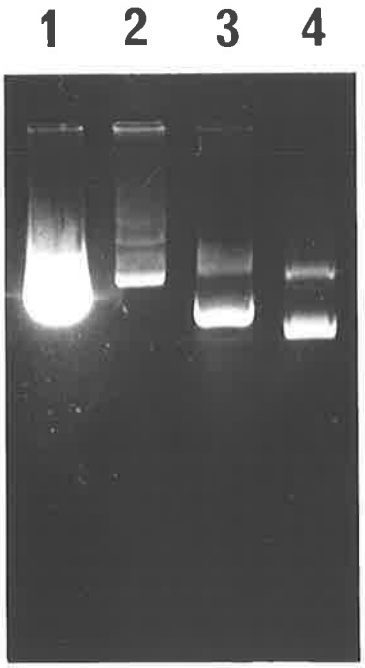
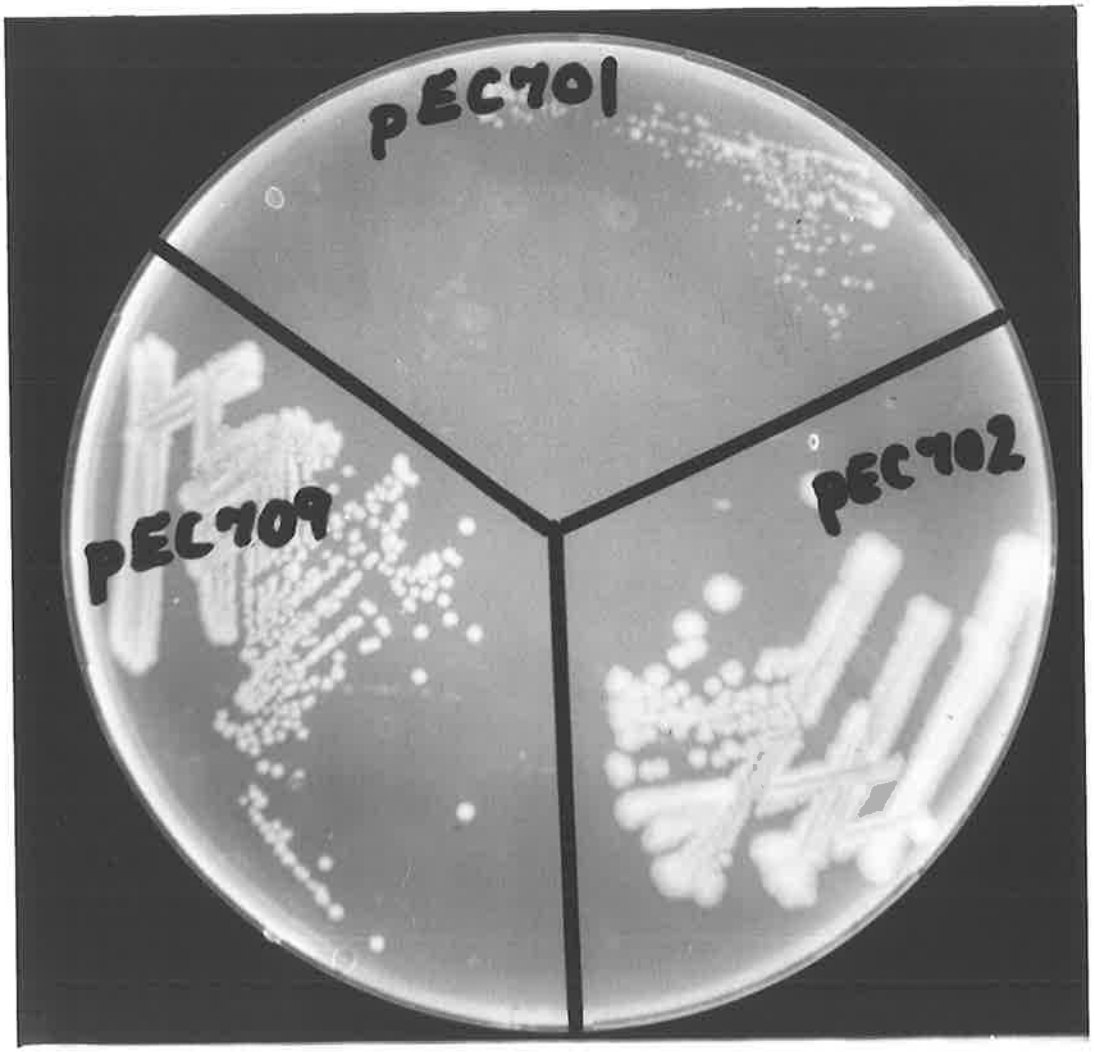

1
2
3
4

Fig. 5.4 Comparison of the colonies of E536 derivatives harboring pEC701, pEC702 and pEC709

Single colonies of the strains E2249 (= E536 (pEC701)), E2250 (= E536 (pEC702)) and E2262 (= E536 (pEC709)) were streaked on YGC + Amp(50 ng/ml) and incubated at 37°C for 30 hours. The plate was, then, photographed under short wave UV.


PEC701
PEC709
PEC702

The normal colony size of E536 derivatives carrying pEC707 also indicated that this plasmid was not replicating from the phage *ori*.

5.3.2.2 Deleting C-terminal end of *RA* prevented replication from 186-ori

Since the reading frame of *RA* extended up to position 4068, deleting the DNA betwen the unique *HpaI* site (position 3941) and *BamHI* (position 4854) site might render the protein non-functional. This deletion was effected by digesting pEC701 DNA with *HpaI* and *NruI** and self-ligating the larger fragment isolated from agarose gel (* the *NruI* site is at 597 bases to the right of *BamHI*, on the pBR part of pEC701; see Fig. 3.4). Plasmid from one colony was tested for its ability to transform *polA*- and found that it lacked this ability. The normal colony character was another evidence to say that replication of the phage *ori* was defective in this plasmid.

The loss of phage-specific replication associated with mutations in *RA* could not be ascribed to a polarity of these mutations on the expression of downstream genes, as *CP93*, the only potential gene downstream to *RA*, was found to be not required for replication (see section 5.3.1).*RA* must therefore be directly needed for replication.

5.3.3 GENES UP TO RA ARE NOT ESSENTIAL FOR 186 REPLICATION

The most direct way to test if *LA* or any other gene to the left of it was needed for replication would be to delete the DNA to the left of *RA*. To avoid the possibility of polar effects on *RA* expression, a ribosome binding site must be fused in phase with the part of the *LA* reading frame which was retained in the insert.

Fig. 5.5 The sequence explaining the construction of pEC706

The fusion of *LA* and *CP95* is depicted in the figure.

A. Partial coding region of *CP95*. (Since the *PstI-BamHI* DNA was cloned into M13mp9, the sequences after the *BamHI* site belong to the vector DNA).

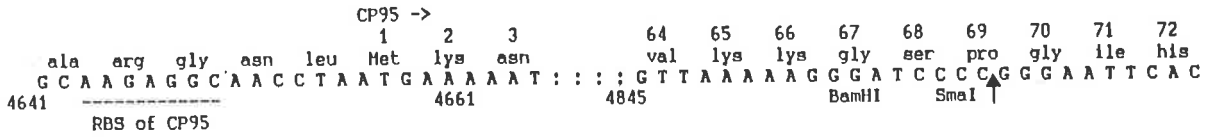
B. Partial coding region of *LA*.

C. Reconstructed sequence after the fusion between *CP95* and *LA*.

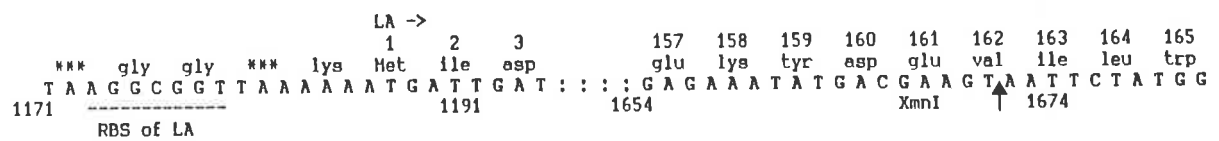
Arrows point to the cleavage sites of the enzymes *SmaI* and *XmnI*.

Sequence numbers corresponding to the bases in the BGLBAM sequence (see Fig. 4.2) are shown unchanged.

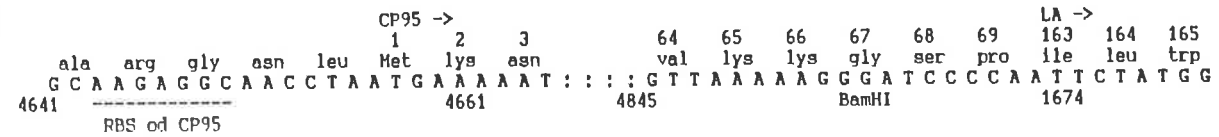
A


CP95 ->
1 2 3
64 65 66 67 68 69 70 71 72
ala arg gly asn leu Met lys asn val lys lys gly ser pro gly ile his
GCAAGAGGCAACCTAATGAAAAAT : : : : GTTAAAAAAGGGATCCCCGGGAATTCAC
4641 4661 4845
BamHI
SmaI
RBS of CP95

B


LA ->
1 2 3
157 158 159 160 161 162 163 164 165
*** gly gly *** lys Met ile asp glu lys tyr asp glu val ile leu trp
TAAGGCGGTTAAAAAAATGATTGAT : : : : GAGAAATATGACGAAGTAATTCTATGG
1171 1191 1654
XmnI
1674
RBS of LA

C


CP95 ->
LA ->
1 2 3
64 65 66 67 68 69 163 164 165
ala arg gly asn leu Met lys asn val lys lys gly ser pro ile leu trp
GCAAGAGGCAACCTAATGAAAAAT : : : : GTTAAAAAAGGGATCCCCAATTCTATGG
4641 4661 4845
BamHI
1674
RBS od CP95

The unique *XmnI* site (position 1666) on BGLBAM fragment is situated in the coding region of *LA* (Fig. 4.2). Furthermore, the *PstI* (4320) to *BamHI* (4854) DNA fragment when blunt-ligated to this site would join the N-terminal coding sequence and the RBS of *CP95* with the reading frame of *LA* (Fig. 5.5). Transcription of the genes could be ensured by providing a promoter in front of this piece of DNA. If this would allow the 186-specific replication, inclusion of a drug-resistance determinant in the making of the recombinant molecule could result in a plasmid which replicated exclusively from 186*ori*. The cam^R gene of pBR329 (Fig. 5.6; and Covarrubias and Bolivar, 1982) appeared to be the best choice for this purpose, as this gene could be isolated, together with its promoter (p^{cam}) but without simultaneously taking the pBR*ori*, by digesting the plasmid DNA with *PstI* and *BamHI*. This could be ligated with the 186 DNA fragments described above to give a molecule which had 186*ori*, *CP95-LA* fusion gene, *RA* and the cam^R gene. It was expected that p^{cam} would transcribe the 186 DNA fragment as well as the cam^R gene. As no pBR*ori* would be present on such molecule, replication of the plasmid, and consequently stable transformation of cells, would indicate the presence of functional 186*ori*.

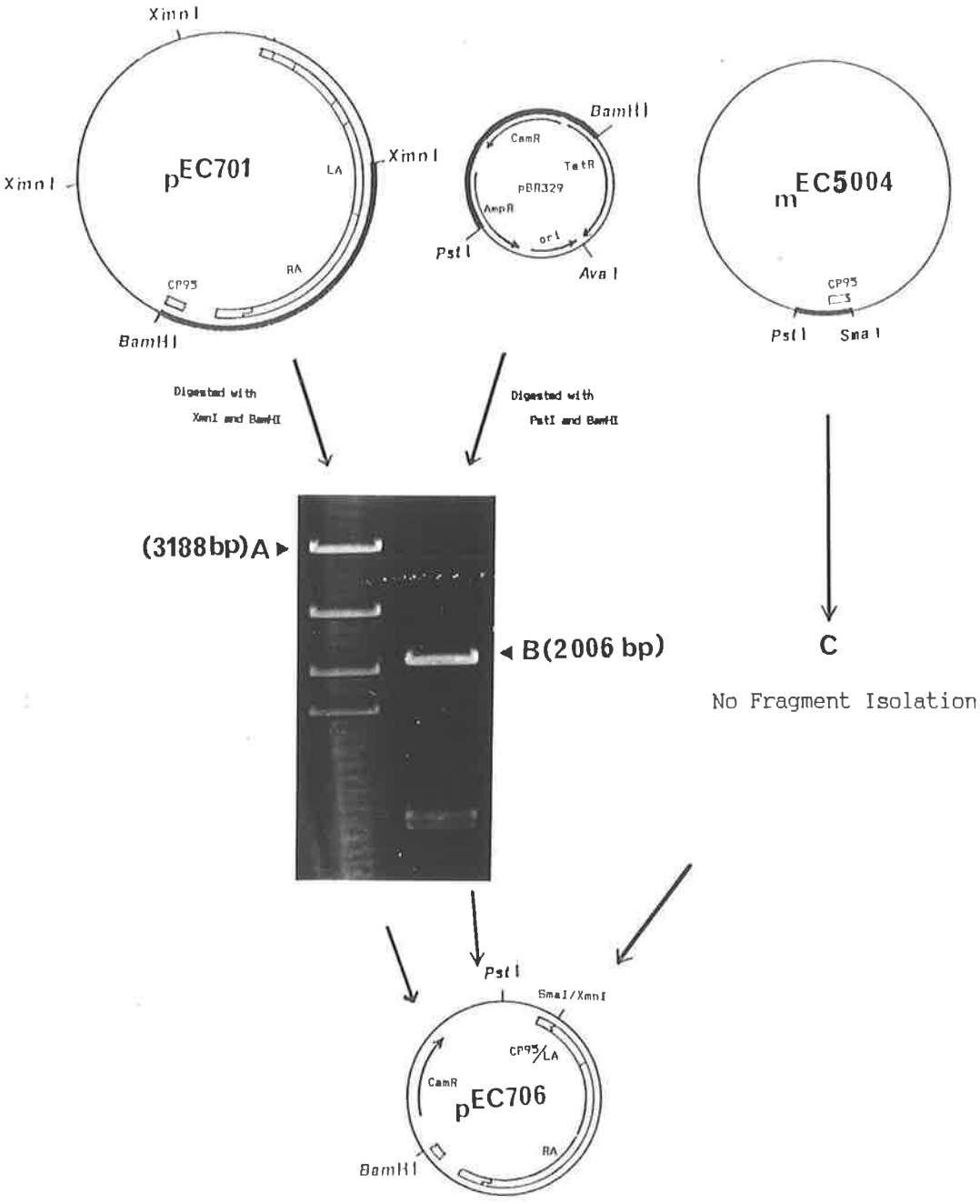
5.3.3.1 Construction of pEC706, a plasmid which replicates exclusively from 186*ori*

A schematic representation of the construction of pEC706 is given in Fig. 5.6. The plasmid pEC701 was cut with *XmnI* and *BamHI* and the 3.2 Kb band, corresponding to 1666 to 4854 of the BGLBAM fragment was isolated from agarose gel (Fragment

Fig. 5.6 Strategy of construction of pEC706

The plasmid pEC701 was digested with *XmnI* and *BamHI* and isolated the largest fragment (A) from 1% agarose gel. The pBR329 DNA was digested with *PstI*, *BamHI* and *AvaI* and the largest fragment (B) was isolated from 1% agarose gel. The RF DNA of mEC5004 (= M13mp9 clone of the *PstI* (4320) to *BamHI* (4854) of the BGLBAM fragment) was digested with *PstI* and *SmaI* and used as such (C).

Fragments A, B and C were ligated together and used to transform E901 for selecting *camR* transformants.


XmnI
pEC701
LA
RA
CP93
BamHI
CamR
TetR
pBR329
AmpR
ori
PstI
AvaI
mEC5004
CP93
PstI
SmaI
Digested with
XmnI and BamHI
Digested with
PstI and BamHI
(3188bp) A
B(2006 bp)
C
No Fragment Isolation
PstI
SmaI/XmnI
CP93/LA
CamR
pEC706
RA
BamHI

A). Due to the approximately equal lengths of the two DNA fragments (2.1 vs 2.0 Kb) which would be produced by digesting pBR329 DNA with *PstI* and *BamHI*, to facilitate the separation of the required band from the other this DNA was further digested with *AvaI* before electrophoresing on 1% agarose gel. The 2.0 Kb *PstI* to *BamHI* fragment was isolated from gel (Fragment B). The DNA fragment containing the RBS of *CP95* (Fragment C) was prepared by digesting the RF DNA of MEC5004, which is an M13 clone of the *PstI* (4320) to *BamHI* (4854) section of the BALBAM fragment with *PstI* and *SmaI*. No fragment isolation was performed in this case. Fragments A, B and C were ligated together and cam^R transformants of E901 were selected. The plasmid DNA from one colony was tested for the *HinfI* digestion pattern and found that the pattern (Fig. 5.7) agreed exactly with the computer prediction, thereby confirming the identity of the new plasmid (pEC706). The structure of this plasmid, as deduced from the sequences of the BGLBAM fragment and of pBR329, is given in Fig. 5.8. Colonies carrying pEC706 were normal sized. Plasmid DNA preparations from stationary phase cultures (Fig. 5.3) did not show the high yield characteristic of pEC701, and therefore it appeared that the copy number of this plasmid was normal.

Since the DNA fragment taken from pBR329 did not contain the plasmid *ori* (Covarrubias and Bolivar, 1982) and therefore replication of the plasmid pEC706 depended on a functional 186*ori*, isolation of self-replicating recombinant molecules of the type given in Fig. 5.8 indicated that none of the genes to the left of *RA* was needed for replication.

Fig. 5.7 Fragments generated by digesting pEC706 with *HinfI*

Plasmid pEC706 DNA was digested with *HinfI*, end-filled in presence of ^{32}P dCTP and electrophoresed on 5% polyacrylamide gel (Ac:Bis 50:1).

Sizes of fragments (bp) were determined from the sequence of pEC706 deduced from that of the BGLBAM fragment and of pBR329 (Covarrubias and Bolivar, 1982).

The *ori* of pBR329 resides in the 396 bp fragment in the track of pBR322 (Sutcliffe, 1979).

pEC pBR
706 322

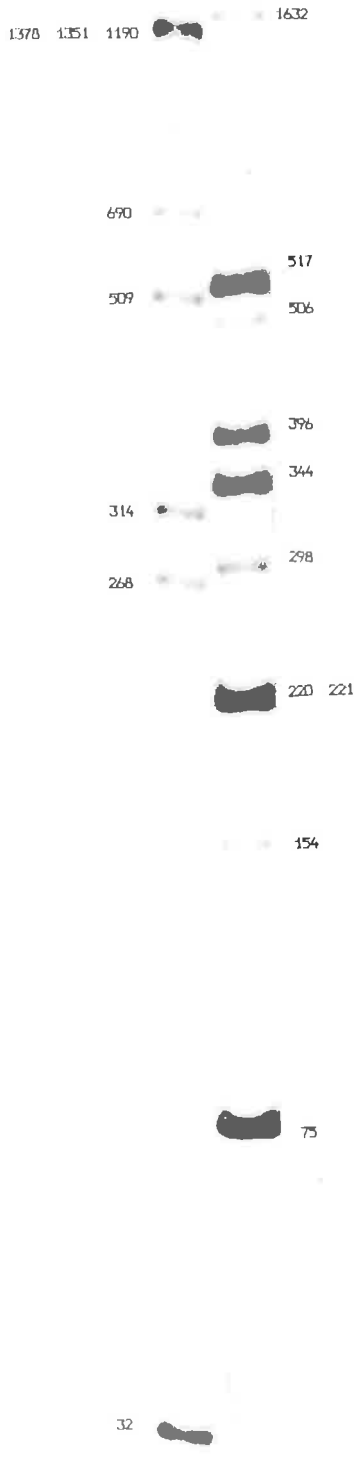
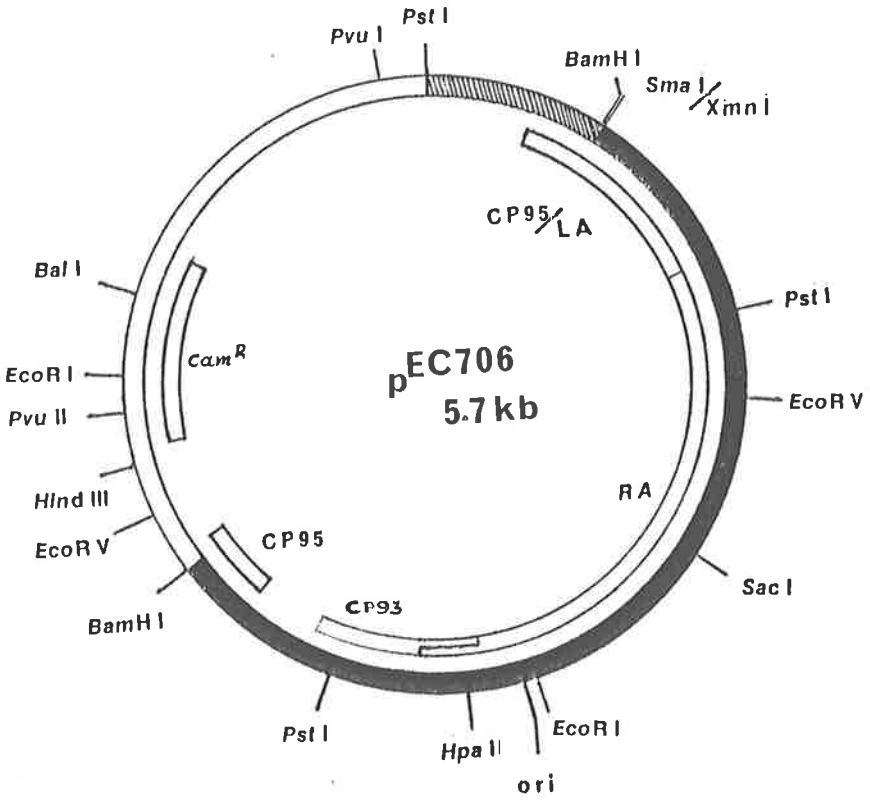


Fig. 5.8 Circular map of 706

The map was prepared from the sequence deduced from the BGLBAM sequence and that of pBR329 (Covarrubias and Bolivar, 1982). The sites of *BamHI*, *EcoRI* and *EcoRV* have been verified by restriction analyses.


Pst I
Pvu I
BamH I
Sma I
Xmn I
CP95
LA
Pst I
EcoR V
RA
Sac I
EcoR I
ori
Hpa II
Pst I
CP93
CP95
BamH I
EcoR V
Hind III
Pvu II
EcoR I
Bal I
cam R
pEC706
5.7 kb

The specificity of replication of pEC706 was further confirmed by its ability to transform the *polA* strain. As expected of 186 replicon this plasmid did not tranform the *rep3* strain.

5.3.4 CONFIRMATION OF THE POLARITY OF Aam11 ON THE EXPRESSION OF RA

The finding that translation of *LA* was needed for the expression of *RA* could be confirmed with the use of pEC706. One of the two *BamHI* sites in this plasmid was located near the junction between *CP95* and *LA* reading frames (Fig. 5.8). Adding four bases to this site would introduce a frameshift in the reading frame of the *CP95-LA* fusion gene, causing the translation initiated from *CP95* to terminate at position corresponding to 1746 of the BGLBAM sequence. This should have similar, if not identical, effect as the *Aam11* mutation on the expression of *RA*. The question could then be asked if the plasmid expressed *RA*, by testing its ability to replicate.

Plasmid pEC706 DNA was cut with *BamHI* and end-filled before ligation. As control, the same DNA stock was cut with *EcoRV*, which also cut pEC706 twice (see Fig. 5.8), and treated with Klenow in presence of all four dNTPs and religated. The protruding ends of *BamHI* digested DNA would be filled in by Klenow, whereas the blunt ends of *EcoRV* (Schildkraut, et. al., 1984) would be unchanged. On religation, these two DNA preparations should give approximately the same number of transformants if the frameshift caused by the end-filling of *BamHI* site was without effect on the expression of *RA*. In fact it was found that the transformability of the *BamHI*

digested DNA was less than 1/1000 of that digested with *EcoRV* (5 c.f.u. with 100 ng of *BamHI* cut DNA versus about 5300 c.f.u. with the *EcoRV* cut DNA). Plasmid DNA from all the five colonies from the former showed the presence of two *BamHI* sites which indicated that they occurred as trace contamination by uncut/unfilled DNA.

The absence of replication caused by frameshift mutation in the reading frame of *LA* gave further evidence to say that the translation of *LA* to its normal termination codon was essential for the expression of *RA*.

Although the results with pEC706 confirmed the need for translation of *LA* for the expression of *RA*, they did not help to distinguish between transcriptional and translational polarities. It however became evident that the product of *LA* was not directly involved either positively or negatively in the regulation of expression of *RA*.

5.3.5 *LA* IS NEEDED FOR THE EFFICIENCY OF PHAGE REPLICATION

Though not directly needed for the replication from 186*ori* the product of *LA* might nevertheless play a role in the efficiency of replication, or in the replication of the whole phage DNA in as yet unknown way. This is somewhat indicated by the fact that pEC706 gives large colonies and normal yield of plasmid DNA. The possibility that *LA* might be needed for the whole phage replication could only be tested if translation of *RA* was made independent of the production of functional *LA* protein. Engineering the reading frame of *LA* in a way similar to what was done in pEC706 was found impractical with whole phage DNA. Besides, involvement of *LA* in some

essential function(s) other than replication would make such phages non-viable. The best alternative available was therefore to isolate a temperature sensitive variant of *LA*. Since an *LAts* phage would be expected to have normal translation to the end of this gene and consequently normal production of *RA* protein, any difference in rate of replication must be due to the absence of functional *LA* protein.

Single step reversion of an *LAam* mutant would be the best way to look for *LAts* phages. Temperature sensitive mutants may result if the amino acid inserted at the site of the amber codon is not acceptable for proper functioning of the protein at high temperature.

Temperature sensitive revertants of 186*cItspAam11* phage were therefore isolated as described in section 5.2.1. Su-(E251) lysogen carrying one such mutant as prophage, together with wild-type prophage control, was used for pulse-labelling (see section 5.2.2) in order to determine the need for LA in phage replication. The peak due to 186 DNA systhesis was absent in the case of *LAts* (Fig. 5.9) which indicated that this protein was needed for the replication of the phage DNA.

Since the product of *LA* was not required for the replication of pEC706, it was assumed that this gene was needed for increasing the efficiency of whole phage replication, or for some facet of replication unique to the whole phage DNA.

Fig. 5.9 Pulse label incorporation into heat induced culture of E4012 (= E251 (186*cItspLAts11*)) and E252 (= E251 (186*cItsp*))

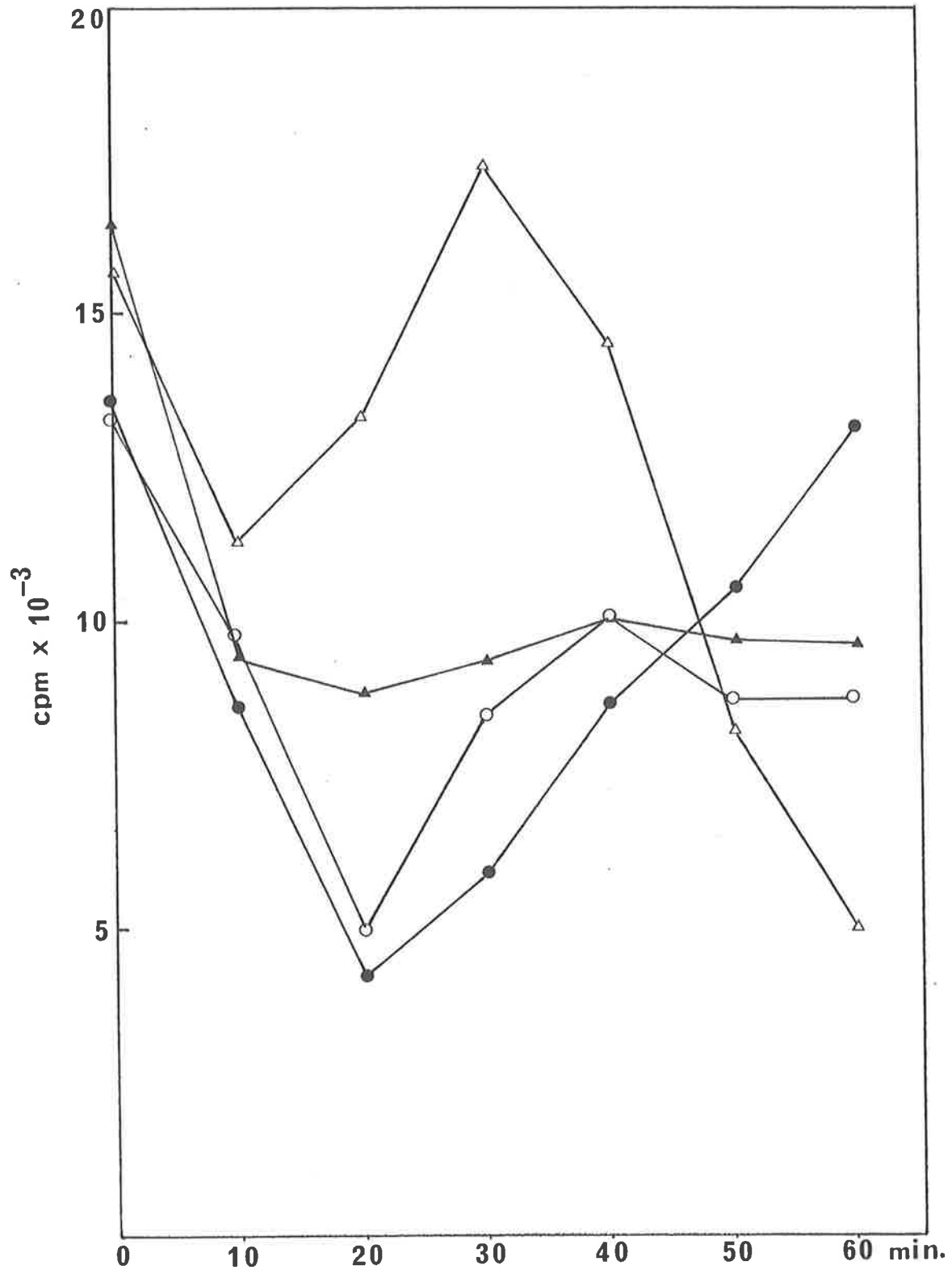
Overnight cultures of E4012 and E252 were prepared in LB at 30°C, sub-cultured into TPGCAA and grown with aeration at 30°C until the A600 reached 0.3. The cultures were then transferred to 40°C and incubation continued with aeration. Pulse-labelling with ^{3}H thymidine was done as described in section 5.2.1.

▲ E4012 at 40°C

△ E252 at 40°C

● E4012, heat induced at 40°C for 10 minutes and, then, transferred to 30°C.

○ E252, heat induced at 40°C for 10 minutes and, then, transferred to 30°C.


20
15
10
5
cpm x 10^{-3}
0
10
20
30
40
50
60 min.

5.4 DISCUSSION

Results reported in this chapter showed that the product of *RA* alone was essential for the replication of plasmids carrying 186*ori*. *LA* was not needed for replication, and the replication deficient phenotype of *LAam* phages could be explained by a polar effect of such mutations of the expression of *RA*. However, a possible role of *LA* in increasing the efficiency of phage DNA replication has been indicated.

CHAPTER 6

GENERAL DISCUSSION

As outlined in chapter 1 the main aim of this thesis was to identify the phage genes essential for the replication of coliphage 186. The approach was to isolate the minimal DNA fragment which was able to exhibit the phage specific replicaiton and, then, characterize the genes and structural features carried on this fragment in order to define the replication genes and *ori*. Besides exposing the phage gene essential for replication, this study also revealed a possible mechanism of control of gene expression for the replication gene.

6.1 MINICHROMOSOME OF 186

Genetic mapping of alleles of *geneA* and physical mapping of *ori* had indicated that these two phage replication functions were contained on the DNA between 79.6% (*BglII* site) and 96% (*BamHI* site) of the chromosome. This section of DNA, termed the BGLBAM fragment, was therefore the initial choice for use in the construction of minichromosome of 186.

Genetic techniques had exposed only *geneA* on this 4.9 kb DNA fragment. The start and stop positions of this gene had been proposed to be 83.8 and 87.0%, respectively, of the chromosome (Finnegan and Egan, 1979). This left about 1.5 kb of DNA between the *BglII* site and the start of *geneA*, and more than 2 kb between the end of this gene and the *BamHI* site, without any known gene function. It was therefore expected that other genes, whose presence had not been detected by analysis of conditional lethal mutants might be encoded by the

BGLBAM fragment. Alternatively, essential genes could escape detection if it was unable to obtain conditional lethal mutations in them. If so, it was possible that the BGLBAM fragment carried essential genes, some of which might even be needed for replication. Identification of such genes could be done with the help of the DNA sequence.

Two alternatives available for the making of minichromosome were to use the BGLBAM fragment to generate a recombinant molecule that replicated exclusively from 186*ori*, or to clone this fragment into a suitable vector and assay the activity of the phage *ori*. The latter alternative was adopted in the first instance by cloning the BGLBAM fragment into the *BamHI* site of pBR322, to generate pEC701. Since the DNA was inserted behind p^{tet}, this method ensured that the genes carried on BGLBAM fragment would be expressed even if no phage promoter was present on the fragment. Actually, absence of phage promoters in between the *BglII* site and replication genes was later confirmed from the DNA sequence. Furthermore, presence of the pBR *ori* on the same moleucle gave the advantage that 186 replication genes carried on the insert could be mutated without loss of replication of the plasmid. Spontaneous and induced mutations that resulted in the absence of replication of 186*ori* could, therefore, be retained and studied.

6.2 EVIDENCE FOR THE 186-SPECIFIC REPLICATION OF pEC701

Ability of pEC701 to transform *polA* strain was indicative of functional 186*ori* on this plasmid. The high yield of plasmid DNA from strains harbouring pEC701 was probably due to

the high copy number of the plasmid, which therefore was supporting evidence for the presence of two functional *ori*. As this high yield was not seen when *RA*, as well as *LA*, amber mutations were present on the plasmid, it indicated that phage-specific replication was responsible for this high yield. The apparently contradicting observation that *log* phase cultures yielded less plasmid DNA could be explained as due to the association of the replicating 186*ori* with cell membrane. Although no direct evidence for such association is available at present, analogy with P2 and Lambda suggests that 186 replication follows the attachment of the phage DNA to the cell membrane. As the plasmid DNA preparation procedures are aimed at selecting supercoiled DNA, molecules which are associated with the cell membrane are likely to be left behind. Assuming that the replication of 186*ori* takes place only in *log* phase cells, this seems to be a likely explanation for the low yield. Since the phage requires several host proteins for its own DNA replication, it is conceivable that in stationary phase cells, which do not support extensive replication, the replication of phage *ori* may be switched off, and the resulting dissociation from cell membrane will then make the isolation of the plasmid DNA molecules possible.

The minute colony character of strains carrying pEC701 could also be ascribed to the phage specific replication of this plasmid, as situations which prevented replication from 186*ori*, such as mutations in *LA*, *RA* and the host *rep* gene, resulted in normal sized colonies.

6.3 MINUTE COLONIES OF pEC701 MAY RESULT FROM TITRATION OF *lexA* REPRESSOR

As mentioned in Chapter 3, the size of colonies carrying pEC701 was always smaller than those carrying pEC702 or the variants of pEC701 which did not exhibit 186-specific replication, and this appeared to be due to slow growth rate of cells. One reason for this reduced growth rate of cells carrying pEC701 was thought to be the simultaneous activity of two origins of replication in this plasmid. However, pEC701 did not produce normal sized colonies in *polA* strains, and this was contrary to what was expected in the absence of replication from the pBR *ori*. This raised the possibility that the replication from 186*ori* was responsible for the minute colony character. Even this did not seem to be the case, as the colonies produced by pEC706 were normal sized. It therefore appeared that some other factor associated with the 186-specific replication of pEC701 was responsible for the formation of the minute colonies. As mentioned in Chapter 3, the yield of pEC701 was significantly higher than that of pEC702 which might represent elevated copy number of this plasmid. Interference to host metabolism caused by this high copy number of pEC701 could be responsible for the slow growth rate of cells, and consequently the minute colonies. It was however found that the yield of pEC709, the plasmid that did not carry the DNA between 4320 to 4854 of BGLBAM fragment, was at the same high level as that of pEC701 but, nevertheless, the colonies were normal sized. This suggested that some function(s) or DNA sequence(s) deleted from pEC701 in the

making of pEC709 was responsible for causing slow growth rate of the cells. Such function(s)/DNA sequence(s) in combination with high copy number of plasmid DNA might have interfered with the host metabolism. Three probable candidates for this are the gene product of CP93, transcription from p95, and the *lexA* binding site (SOS Box) situated at 4482 of the BGLBAM sequence. With the available data it is not possible to rule out any of these three possibilities but the following model based on the SOS box seems to be the most logical explanation:

Multiple copies of the SOS box can lead to a titration of the *lexA* repressor of the host. As the *lexA* repressor is involved in the regulation of the *E. coli* gene which causes filamentation (*sfi*) (George *et. al.*, 1975) titration of this repressor by the replicating pEC701 can possibly lead to derepression of this gene. This might be responsible for the slow growth rate and the accompanying minute colony character of strains carrying pEC701. One way of testing this hypothesis would be to see the colony morphology of the *sfi* strains harbouring pEC701.

The plasmid pEC706, which replicated from 186*ori* and carried the SOS box, did not result in minute colonies. This was probably due to the relatively low copy number at which this plasmid was maintained as compared with the high copy number of pEC701.

6.4 CHARACTERIZATION OF THE BGLBAM FRAGMENT

The first step towards identification of the replication genes was the determination of the DNA sequence of BGLBAM fragment. This was done by the dideoxy chain termination method and the complete nucleotide sequence was determined from both strands of DNA. Presence of six potential genes transcribed rightwards on the BGLBAM fragment was revealed by the DNA sequence, and later the minicell studies indicated the reality of all but one gene. Besides exposing these potential replication genes, the DNA sequence also revealed sites for several restriction enzymes which were later used in pruning down the BGLBAM fragment in order to identify the gene essential for replication.

6.4.1. *GeneA* is exposed as *LA* and *RA*

The most prominent feature emerged from DNA sequencing was that *geneA*, all amber alleles of which had been grouped into one complementation group was exposed as two separate genes, *LA* and *RA*. Authenticity of this finding was confirmed by sequence mapping of *Aamber* alleles in the reading frames of *LA* and *RA* as well as by direct visualization of protein bands corresponding to these genes. Absence of complementation between alleles of *LA* and *RA* could either be due to the *cis*-action of both polypeptides or due to the *cis*-dominant dependence of expression of one gene on the other. The latter seemed to be the case as the *LAam11* mutation was found to prevent the expression of *RA*.

The potential genes *CP80*, *CP81*, *CP83* and *CP93* might have escaped detection during the genetic analysis of amber

mutations, either due to their possibly non-essential gene products or due to some biological constraints in obtaining conditional lethal mutations in them.

6.4.2 Identification of the replication gene

Of all the genes carried on the BGLBAM fragment, *RA* appeared to be the most likely candidate for being one replication gene as its coding region overlapped the EM *ori*. Replication genes usually overlap the *ori* as evidenced by the overlapping of *geneO* of lambda (Schnos and Inman, 1970; Inman, 1981; Tsurimoto and Matsubara, 1983), *geneA* of phiX174 (Langveld *et. al.*, 1979; Francke and Ray, 1972; Eisenberg *et. al.*, 1977; Geisselsoder, 1976), and *geneA* of P2 (Schnos and Inman, 1971), with the respective origins of these phages. Since these genes are involved in the initiation of phage DNA replication, analogy with them therefore suggested that *RA* might be involved in the actual initiation of replication. The participation of *RA* in replication of pEC701 was indicated by the inability of plasmids to transform *polA* strain when this gene was mutated.

The strategy adopted was therefore to delete from pEC701 the genes on either side of *RA* and ask if the plasmid still retained 186-specific replication.

It was found that deleting the C-terminal end of *CP93* had no effect on replication and so this gene, if genuine, did not take part in replication. It could however be argued that the loss of 18 amino acids from the C-terminal end of this gene did not abolish the activity of the protein and so the plasmid was capable of replicating from the phage *ori*. Due to the

overlapping of the reading frames of the genes *RA* and *CP93* it was not possible to delete the entire coding region of *CP93* without simultaneously affecting the expression of *RA*. Any other method of mutagenesis in the non-overlapping part of *CP93* would not help to rule out the above mentioned argument.

Deleting *LA*, and the other genes to the left of it, also did not prevent 186-specific replication of pEC701, but it was found that providing facility for the delivery of ribosomes to the start site of *RA* was essential for replication. A possible cause and significance of this dependence of *RA* on delivery of ribosomes is discussed later.

The conclusion therefore was that *RA* alone was essential for the replication of 186*ori*. Since its expression depended on the delivery of ribosomes through the reading frame of *LA*, it was indirectly under the control of the latter. Absence of complementation between amber alleles of these genes was therefore probably due to this *cis*-dominant effect.

6.5 ROLE OF LA IN REPLICATION

Though not essential for replication initiated from the 186 ori, the product of *LA* might be involved in increasing the efficiency of replication. This was initially indicated by the finding that the yield of pEC706, where part of *LA* is deleted, was less than that of pEC701. When *LAts* phage was used for pulse labelling studies to determine phage replication it was found that the peak in label incorporation, characteristic of 186 replication, was absent at the high

temperature. The *LAts* phage, nevertheless, was able to produce plaques at high temperature, but these plaques were very minute as compared with the *LA*+ plaques. The appearance of plaques could be attributed either to an incomplete inactivation of the *LA* gene product or, more likely, due to some replication effected by the product of *RA*. A working model for 186 replication therefore proposes that *RA* alone is essential for the initiation and subsequent elongation of DNA chain from the phage *ori*, but the *LA* was needed for achieving high rate of replication. This effect of *LA* could be either at the level of initiation or at elongation. A gene analog of *LA* has not been identified in phages lambda, P2 or any other coliphage and so *LA* and its role in increasing the efficiency of replication are worth studying. One hypothetical role that can be proposed for this gene is in the multiple initiation from 186*ori*. As mentioned in Chapter 1, phage 186 is capable of repeated initiations from its *ori*. This character has not been found for phages lambda and P2 and so represents a 186-specific character. *LA* could be envisaged as a gene whose product enables the phage to repeatedly initiate from its *ori*.

6.6 CONTROL OF EXPRESSION OF RA

The disappearance of the *RA* protein band in the presence of *LAam11* mutation suggested a transcriptional polarity of this amber mutation. Alternatively, the product of *LA* might have been directly controlling the expression of *RA*. This possibility was however ruled out by showing that deleting

part of *LA* without affecting the delivery of ribosomes to the start site of *RA* did not result in the absence of 186-specific replication. It therefore appeared that transcriptional polarity or translational coupling of these genes was responsible for this dependence of *RA* on translation of *LA*. The work described in this thesis does not help to distinguish between these two possibilities but it seems appropriate to closely examine the DNA sequence near the start site of *RA*, as it does not show the presence of a suitable ribosome binding site (RBS) (Figs. 6.1 and 4.2). The reading frames of these two genes overlap in the tetra nucleotide sequence ATGA (Fig. 4.2). Due to this arrangement of the reading frames, the ribosomes which translate *LA* will terminate near the initiator codon of *RA*. In other reported cases of translational coupling, the delivery of ribosomes to the vicinity of the initiator codon has been found to increase the efficient translation of the distal gene (Schumperli *et. al.*, 1982). The translationally coupled *trpE* and *trpD* genes (Oppenheim and Yanofsky, 1980; Aksoy *et. al.*, 1984) and *trpB* and *trpA* (Aksoy *et. al.*, 1984) show overlapping termination and initiation codons. In *galT-galK* coupling, the termination codon of the *galT* gene is placed only three bases away from the start of the *galK* (Schumperli *et. al.*, 1982; Normark *et. al.*, 1983). Two alternative mechanisms have been proposed for the coupling in situations like this (Schumperli *et. al.*, 1982). In the first case the localized concentration of ribosomal subunits in the vicinity of the initiator codon, caused by the dissociation of ribosomes from the mRNA on encountering the stop codon, is needed for the binding of ribosomes to the

Fig. 6.1 Sequence showing the area around the beginning of *RA*

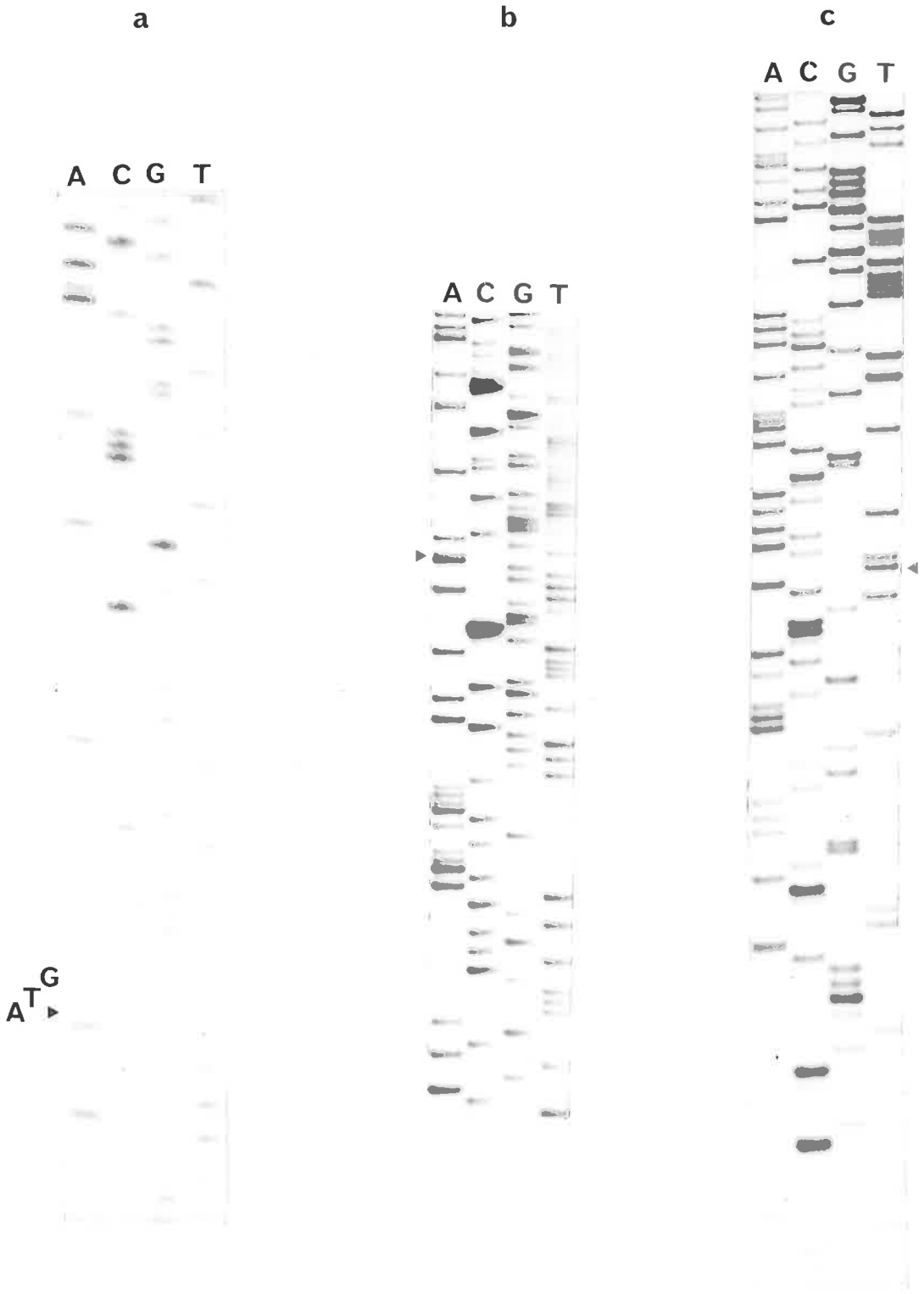
The sequences were determined with the universal primer on M13mp9 clones of *HpaII* or *AluI* fragments.

a. Sequence of the *HpaII* fragment from 2174 to 2586, on the *l*-strand.

b. Sequence of the *AluI* fragment from 2076 to 2241, on the *l*-strand.

c. Sequence of the *AluI* fragment from 2076 to 2241, on the *r*-strand.

The start codon of *RA* is marked by the arrows.

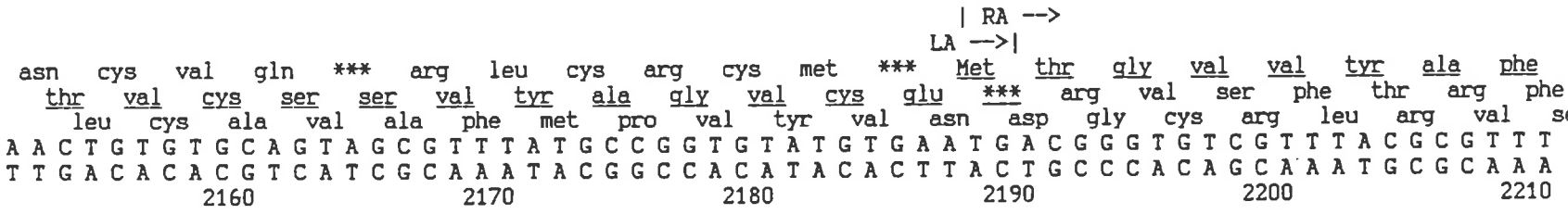

a
A C G T
A T G
b
A C G T
c
A C G T

ribosome binding site of the distal gene. In the second possibility, the ribosomes which terminate translation of the proximal gene may not dissociate from the mRNA but instead continue translation from the next available sense codon. This explanation fits well with the situation in *galT-galK* coupling where the two reading frames are in phase (Schumperli *et. al.*, 1982) whereas in the case of *LA* and *RA* the reading frames are different and therefore a shift in the reading frame will be essential for the ribosome to continue translation in the frame of *RA*. Presence of an in-frame nonsense codon immediately preceding the initiator codon (Fig. 6.2) of *RA* will make the translation to stop before encountering the start site of this gene, even if frameshift occurs during the translation of *LA*. The most probable means for coupling between these two genes is therefore the increased localized concentration of ribosomal subunits near the vicinity of the initiator codon of *RA*. Presumably, the high concentration of the ribosome forming units helps in the binding of ribosomes to the start site of this gene, even though no detectable RBS precedes its initiator codon.

Dependence of a gene's expression on frameshift error occurred during the translation of the gene preceding it has been exemplified by the single-strand RNA phage MS2. In this phage the translation of the lysis gene depends on the delivery of ribosomes to the vicinity of its initiator codon by a frameshift error during the translation of the coat gene which precedes it. (Kastelein, *et. al.*, 1982; Kastelein and van Duin, 1982)

Fig. 6.2 Sequence showing the *LA-RA* overlap

The reading frames of *LA* and *RA* are underlined.


| RA -->
LA -->|
asn cys val gln *** arg leu cys arg cys met *** Met thr gly val val tyr ala phe
thr val cys ser ser val tyr ala gly val cys glu *** arg val ser phe thr arg phe
leu cys ala val ala phe met pro val tyr val asn asp gly cys arg leu arg val se
A A C T G T G T G C A G T A G C G T T T A T G C C G G T G T A T G T G A A T G A C G G G T G T C G T T T A C G C G T T T
T T G A C A C A C G T C A T C G C A A A T A C G G C C A C A T A C A C T T A C T G C C C A C A G C A A A T G C G C A A A
2160 2170 2180 2190 2200 2210

Yet another way in which upstream translation helps in the translation of downstream gene is by ironing out the mRNA secondary structures which block the access of ribosomes to the ribosome binding site of the downstream gene (cited by Normark *et. al.*, 1983). The absence of ribosome binding site for *RA* makes this explanation inapplicable in the case of *LA-RA* coupling.

The involvement of *LA* in replication of whole phage suggests a significance for the coupling between this gene and *RA*. As was proved by the ability of plasmids to replicate using only the product of *RA*, this gene alone might be able to replicate the whole phage DNA. The efficiency of replication to produce enough number of phage particles may, however, depend on functional *LA* protein. If so, dependence of *RA* on the translation of *LA* could mean that the phage was trying to ensure that *LA* protein was produced before replication commenced.

The low level of *RA* production may also help to explain the *in vivo* observation that transcription to the right of *PstI* site at 87.5% was lower than that to the left of this site (Finnegan and Egan, 1981). As mentioned in Chapter 1, a potential stem loop structure in the region of 2270 can act as a transcription terminator. Since the presence of translating ribosomes can prevent the formation of stem loop structures (Yanofsy, 1981) translation of *RA* can 'save' its own message from premature termination by not allowing this stem loop to form. However, if the rate of initiation of *RA* is lower than the rate of transcription, then most of the message molecules will be free of ribosomes and therefore prone to termination

at the stem loop structure. The significance for this is not clear, but a possibility exists that the termination of transcription in this region helps to minimize the transcription across the *ori* region. The DNA sequence does not indicate the presence of any promoter in the region between the two *PstI* sites located at 2362 and 4320 of the sequence. As no gene other than *RA* is encoded by the DNA in this area, presence of another independent transcript originating between these sites seems unlikely.

6.7 OVERLAPPING GENES

An interesting feature of the BGLBAM sequence was the organization of genes on this DNA fragment. All genes from *CP79* to *RA* formed a chain of interlinked genes by having their respective termination codons overlapping with the initiation codon of the next. Except for *CP79/CP80* this overlap was by sharing the sequence ATGA between adjacent genes, and in that respect resembled the overlapping observed in the *nin* region of phage lambda (Kroger and Hobom, 1982; Sanger *et. al.*, 1982). Overlapping termination and initiation codons have been found for the *E. coli* operons such as *his* (Barnes and Tuley, 1983), *frd* (Cole, *et. al.*, 1982), *tox* (Yamamoto *et. al.*, 1982) and *trp* operons (Schumperli *et. al.*, 1982; Aksoy *et. al.*, 1984). Among bacteriophages, phiX174 (Sanger *et. al.*, 1977; Barrel *et. al.*, 1976), G4 (Godson *et. al.*, 1978) and T7 (Dunn *et. al.*, 1981) show overlapping between termination and initiation codons. In addition to aiding in the

translational coupling, such overlapping arrangement can lead to co-ordinate expression of the genes (Normark *et. al.*, 1983) which is a likely explanation for their occurrence in operons and other gene clusters. The BGLBAM fragment represents an early region of the phage chromosome, and the genes carried on this DNA therefore are probably needed for early functions. Since the genes *CP79* to *LA* occurred before *RA* it seemed likely that these genes were expressed prior to, or together with, the expression of *RA*. Except *LA*, no other gene has been indicated to have even indirect role in replication. Significance of their co-ordinate expression with replication gene is, therefore, unclear.

6.8 STRUCTURAL FEATURES

6.8.1 Origin of replication

At present, the position of the *ori* consensus CACTAT is considered as the putative origin of 186. This six-base consensus sequence has been found in or near the origins of lambda (Hobom *et. al.*, 1979), M13 (Meyer *et. al.*, 1979), phiX174 (Sanger *et. al.*, 1977; Langveld *et. al.*, 1978) and T7 (Fuller *et. al.*, 1983). Locating this sequence at 92.0% of 186 chromosome fits well with the determined position of *ori* (92.9 ± 1.8%; Chattoraj and Inman, 1973). The finding that the single strand nick which marks the initiation of rolling circle replication in phiX174 and M13 are only 10 and 1 base, respectively, from this consensus sequence suggests that the nick associated with 186 replication may be near to this site

found at 92.0% of the BGLBAM sequence. Another occurrence of the sequence CACTAT have been found on the *r*-strand of the BGLBAM fragment (position 936 from the *BglII* end) but this site is outside the *ori* limits and so it probably occurred by chance alone. Attempts to determine the exact site of the singe-strand specific nick that initiated rolling circle replication were unsuccessful. Reliance on *in vivo* replication was perhaps the reason for the failure to obtain nicked strands of pEC701 DNA. Successful *in vitro* replication of the phage DNA might help to identify the exact location of the origin of replication.

Unlike in lambda, the *ori* region of 186 did not show sequence repeats. Potential to form secondary structures was also very low in this region. In this respect the *ori* resembled that of phiX174 which also did not have extensive secondary structures (Sanger *et. al.*, 1977). No similarity, however, was seen at the DNA sequence level between these phages.

The DNA sequence near *oriC* shows an unusually high occurrence of the sequence 'GATC' (Kornberg, 1982) whereas no tetra or penta nucleotide sequence was seen at unusually high frequency in or near the 186 *ori*. Interestingly the sequence 'GATC' occurred at a very low frequency in the whole of BGLBAM sequence (twice only).

6.8.2 IHF Box

The consensus sequence for the binding of the Integration Host Factor (IHF; Craig and Nash, unpublished) was found at the beginning of *LA*. This host protein is known to regulate

gene expression at the level of transcription (Friedman *et. al.*, 1984) or at the level of translation (Hoyt *et. al.*, 1982). In many instances the action of this protein is through specific binding to the intercistronic regions of genes as well as to promoters (Freundlich, Friden and Tsui, unpublished). Finding this sequence at the start of *LA* was therefore suggestive of a possible control of expression by IHF binding. Since the translation of *RA* is directly linked to translation of *LA*, the IHF binding can control both genes simultaneously. Role for this possible control of expression of *LA* and *RA* is not known.

6.9 PROSPECTS WITH pEC701

As the minute colony character associated with pEC701 was a result of replication initiated from the phage *ori*, normal colony variants might be able to reveal mutations that prevented 186 replication.

At least three kinds of normal colony variants can be visualized.

Firstly, they could result from plasmid carried mutations which prevented expression of *RA* and/or initiation from 186*ori*. Such colonies could be distinguished from others by the fact that the normal size character would be transmissible with the plasmid. The revertants obtained from pEC701 transformants obviously fell into this class.

In the second class of normal colony variants the mutations residing on the host chromosome might be responsible for

the loss of 186-specific replication of the plasmid. This class of mutations would be of advantage in the study of the host functions required for the replication of phage 186. The host mutations exposed in this way would be in genes which were non-essential for the survival of the cell but essential for the replication of phage 186. Existence of such genes is suggested by the availability of the *rep* mutations. Identification of such non-essential replication genes would be advantageous not only for the study of replication of the phage but also for a better understanding of replication in *E. coli* itself.

In the third class, however, the mutation that conferred normal colony size could be found in either the *lexA* gene itself or in any of the potentially lethal host genes regulated by this repressor. The *lexA* controlled genes fall in the broad catagory of SOS genes, which are induced by DNA damaging agents such as UV, Nalidixic acid, Coumermycin etc. Identification of such these genes might be of significance in the understanding of induction of SOS functions.

6.10 PROSPECTS WITH pEC706

Plasmid pEC706 did not carry *LA* and so it could be used in conjunction with pEC701 to study the possible roles of this gene in replication.

This plasmid could also be used in the study of various host genes involved in replication of phage 186. Phage 186 is known to require the products of *dnaA* and *dnaC* genes of

E. coli (Hooper and Egan, 1981). The plasmid pEC706 also was found to require the product of *dnaA* for replication. Since the amount of genetic information contained on this plasmid is small as compared with the whole phage DNA, the study of involvement of the *dnaA* gene in phage replication would be facilitated by its use.

Another potential use of the plasmid pEC706 would be in developing it as a vector for cloning foreign DNA. Since it carries a replicon compatible with pBR replicon, pEC706 could be used to introduce DNA fragments into strains already carrying pBR plasmids and clones. This would be especially useful when needed to study the *cis/trans* actions of genes.

6.11 FUTURE WORKS

6.11.1 Control of expression of *RA*

One significant finding from the present study was that *geneA*, which was once thought to be a single gene, was exposed as *LA* and *RA*. Of these, only the product of *RA* was essential for the phage replication whereas that of *LA* was needed for the efficiency of replication, and its translation was needed for the expression of *RA*. Control of expression of this essential replication gene, *RA*, through a possible translational coupling with *LA* is worth studying. Firstly, however, the possibility of transcriptional polarity must be ruled out. The most direct way to test this is to detect the transcription patterns in wild type and *LAam* phages. Alternatively, a ribosome binding site can be attached

directly to the initiator codon of *RA*, and assay the effect of *LAam* mutations on the expression of this gene.

Another *in vivo* test to distinguish between transcriptional polarity and translational coupling is to fuse a *lacZ* (β-galactosidase) gene in phase with the reading frame of *RA*. The *lacZ* gene can be used with and without its ribosome binding site, and the effect of frameshift (or nonsense) mutations in the coding region of *LA* on the expression of β-galactosidase monitored. In the event of transcriptional polarity, the expression of β-galactosidase will be affected whether or not the *lacZ* gene possesses its own ribosom binding site. If, on the other hand, translational coupling was the only reason for the dependence of the expression of *RA* on the translation of *LA*, the presence of RBS for the *lacZ* gene will make its expression independent of translation of the *LA* reading frame.

6.11.2 Role of LA in Multiple initiation

The exact role of *LA* in increasing the efficiency of replication is another aspect to be studied further. One foreseeable role of *LA* is in effecting multiple initiations from the phage *ori*. Plasmid pEC701 and pEC706 obviously differ in the efficiencies of replication as evidenced by the difference in yields of plasmid DNA. It must be found out whether the efficiency of replication of pEC706 can be increased by providing the product of *LA*. In fact, supplying *LA* gene product in *trans* was without effect in increasing the yield of pEC706 (unreported observation) and so the experiment must be done by providing this gene product in *cis*. Either a DNA fragment in which *LA* and *RA* are arranged in the normal way

as in pEC701, or one in which the spatial arrangement has been altered must be used to construct a pEC706-type plasmid and compare its efficiency of replication with that of pEC701.

A direct visualization of the replicating molecules of pEC701 and pEC706 under electron microscope may help to distinguish between the differences in initiations from these two plasmids. Alternatively, the *LAts* phage can be used to find out if multiple initiations are absent in this phage.

6.11.3 Localization of the exact origin of replication

Locating the exact origin of replication is of significance in understanding the replication of this phage. Attempts to locate the nick at *ori* by using pEC701 in *rep3* strain was unsuccessful. Presence of functional pBR *ori* on the same moleucle might have contributed to this failure. As pEC706 has only the phage *ori* it must be possible to use this plasmid to locate the nick. One experimental approach is to synchronise the replication by holding a *dnaA* strain carrying this plasmid at the non-permissive temperature and, after lowering the temperature give pulses of label, followed by extraction and restriction analysis of the plasmid DNA. A foreseeable handicap of this scheme is that the pulse has to be extremely short but effective in order to avoid the replication fork traversing the entire length of the molecule during the pulsing. Assuming that the replication rate was the same as that of the *E. coli*, replication of the entire plasmid DNA (about 5 kb) would take only about 3 seconds!

6.11.4 Control of transcription from p95

(a) By the *lexA* binding

Control of expression of p95 through the binding of *lexA* repressor can perhaps be verified by adding purified *lexA* repressor protein into the *in vitro* transcription system. DNA protection studies using *lexA* repressor will tell whether this protein actually binds to the -10 region of p95. A similar experiment using integration host factor will tell whether the consensus sequence (IHF box) found near the start of *LA* actually binds this factor.

(b) By the *dnaA* binding

Presence of the *dnaA* binding site near the start of the message initiated from p95 must also be confirmed by DNA-protein binding studies.

FINAL COMMENT

This thesis exposed *geneA* of phage 186 as two separate genes, *LA* and *RA*, which probably would not have been possible with genetic techniques alone. It has been shown that the phage requires only the product of *RA* for initiating replication from its *ori* whereas the product of *LA* is needed for efficient replication.

The value of DNA sequencing in the characterization of genes has been demonstrated in this work with the following findings:

Firstly, the control of *RA* through its overlapping with *LA* was revealed by the sequence.

Presence of other potential genes in the early region of 186 chromosome has been indicated by the DNA sequence, and therefore study of their functions can now be pursued. The overlapping of these genes is interesting because of their potential co-ordinate expression as well as because it draws some similarity between 186 and lambda as far as the organization of genes is concerned.

Presence of the *lexA* binding site on the -10 region of p95 was an unexpected finding from the sequence, and the potential role of this site in the regulation of expression of this promoter can now be studied further. Similarly, the IHF and *dnaA* binding sites found on the sequence too were outcomes which were not anticipated. Possible roles of these sites in the regulation of gene expression in 186 can be visualized.

Lastly, but not the least, the 186 minichromosome (pEC701), with the minute colony character associated with its phage-specific replication, provided an excellent tool for the isolation and study of the host genes involved in the replication of this phage and therefore was a bonus from this work.

REFERENCES

Aksoy, S. C.L. Squires and C. Squires (1984) *J. Bacteriol.* **157**, 363-367.

Atlung, T. (1981) *ICN-UCLA Symp. Mol. Cell. Biol.* 22, 297-314.

Austin, S. (1976) *J. Bacteriol.* **127**, 32-39.

Bachmann, B.J. and K.B. Low (1980) *Microbiol. Rev.***44**, 1-56.

Bachmann, B.J. (1983) *Microbiol. Rev.* **47**, 180-230.

Bagdasarian, M.M., M. Izakowska and M. Bagdasarian (1977) *J. Bacteriol.* **130**, 577-582.

Bagdasarian, M.M., M. Izakowska, R. Natorff and M. Bagdasarian (1978) *DNA Synthesis* Ed. Molineux, I. and M. Kohiyama, Plenum Press, New York and London.

Baldwin R.L., P. Barrand, A. Fritsch, D.A. Goldthwait and F. Jacob (1966) *J. Mol. Biol.*, **17** 343-357.

Barnes, W.M. and E. Tuley (1983) *J. Mol. Biol.* **165**, 443-459.

Barrel, B.G., G.M. Air, C.A. Hutchison III (1976) *Nature (London)* **264**, 34-40.

Bastia, D., N. Sueota and E. Cox (1975) *J. Mol. Biol.***98**, 305-320.

Bertani, G. (1975) *Mol. Gen. Genet.* **136**, 107-137.

Bertani, L.E., G. Bertani (1971) *Adv. Gen.* **16**, 199-237.

Bird, R.E., J. Louarn, J. Martuscelli and L. Caro (1972) *J. Mol. Biol.* **70**, 549-566.

Birnboim, H.C. and J. Doly (1979) *Nucleic Acids Res.*, 7, 1513.

Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, H.W. Boyer (1977) *Gene* 2, 95-113.

Bowden, D.W., R.S. Twersky and R. Calendar (1975) *J. Bacteriol* **124**, 157-175.

Bradley, C., O.P. Ling and J.B. Egan (1975) *Mol. Gen. Genet.* **140**, 123-135.

Brent, R. and M. Ptashne (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4204-4208.

Calendar, R., B. Lindqvist, G. Sironi and A.J. Clark, (1970) *Virology* **40**, 72-83.

Carl, P. (1970) *Mol. Gen. Genet.* **109**, 107-122.

Chattoraj, D.K. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1685-1689.

Chattoraj, D.K., M. Schnos, R.B. Inman (1973) *Virology* **55**, 439-444.

Chattoraj, D.K. and R.B. Inman (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1768-1771.

Chen, H.R., M.O. Dayhoff, W.C. Barker, L.T. Hunt, L.-S. Yeh, D.G. George and B.C. Orcutt (1982) *DNA* **1**, 365-374.

Cole, S.T., T. Grundstrom, B. Jaurin, J.J. Robinson and J.H. Weiner (1982) *Eur. J. Biochem.* **126**, 211-216.

Cooper, P.K. and P.C. Hanawalt, (1972) *Proc. Natl. Acad. Sci U.S.A.* **69**, 1156.

Covarrubias, L. and F. Bolivar (1982) *Gene* **17**, 79-89.

Davis, R.W., D. Botstein and J.R. Roth (1980). "Advanced Bacterial Genetics." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Delius, H., C. Howe and A.W. Kozinski (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 3049-3053.

Delucia, P. and J. Cairns (1969) *Nature* **224**, 1164-1166.

Denhardt, D.T., M. Iwaya and L.L. Larison (1972) *Virology* **49**, 486-496.

Doudney, C.O. and D. Billen (1961) *Nature* **190**, 545-546.

Dunn, J., M. Elzinga, K.K. Mark and F.W. Studier (1981) *J. Biol. Chem.* **256**, 2579-2585.

Efstratiadis, A., J.N. Vournakis, H. Donis-Keller, G. Chaconas, D.K. Dougall and F.C. Kafatos (1977) *Nucl. Acids Res.* **4**, 4165-4174.

Eisenberg, S., J. Griffith and A. Kornberg (1977) *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 3198-3203.

Finnegan, J. and J.B. Egan (1979) *Mol. Gen. Genet.* **172**, 287-293.

Finnegan, J. and J.B. Egan (1981) *J. Virol.* **38**, 987-995.

Francke, B. and D.S. Ray (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 475-479.

Friedman, D.I., E.J. Olson, D. Carver and M. Gellert (1984) *J. Bacteriol.* **157**, 484-489.

Fuhrman, S.A., D.L. Deininger, P. La Porte, T. Friedmann and Geiduschek, E.P. (1981) *Nucleic Acids Res.* **9**, 6439-6456.

Fuller, R.S. and A. Kornberg (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 15817-5821.

Fuller, C.W., B.B. Beauchamp, M.J. Engler, R.L. Lechner, S.W. Matson, S. Tabor, J.H. White and C.C. Richardson (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 669-679.

Funnell, B.E. and R.B. Inman (1983) *J. Mol. Biol.* **167**, 311-334.

Furth, M.E., C. McLeester and W.F. Dove (1978) *J. Mol. Biol.* **126**, 195-225.

Furth, M.E. and S.H. Wickner (1983) *"Lambda II" Cold Spring Harbour Laboratory* p. 145-173.

Geisselsoder, J. (1976) *J. Mol. Biol.* **100**, 13-22.

Giphart-Gassler, M., J.N. Reeve and P. van de Putte (1981) *J. Mol. Biol.* **145**, 165-191.

Godson, G.N., B.G. Barrel, R. Staden and J.C. Fiddes (1978) *Nature* **276**, 236-247.

Gonzalez, N., J. Wiggs and M.J. Chamberlin (1977) *Arch. Biochem. Biophys* **182**, 404-408.

Gribskov, M., J. Devereux and R.R. Burgess (1984) *Nucleic Acids Res.* **12**, 539-549.

Gromchedl, R. and G. Hobom (1979) *Nature (London)* **227**, 621-627.

Grosveld, F.G., H.-H.M. Dahl, E. de Boer and R.A. Flavell (1981) *Gene* **13**, 227-237.

Hansen, E.B., F.G. Hansen and K. von Meyenburg (1982) *Nucleic Acids Res.* **10**, 7373-7385.

Hansen, F.G. and K. von Meyenburg (1979) *Mol. Gen. Genet.* **175**, 135-144.

Hendrickson, W., H. Yamaki, J. Murchie, M. King, D. Boyd and M. Schaechter (1981) *ICN - UCLA Sym. Mol. Cell Biol.* **22**, 79-90.

Henry, J.J. and R. Knippers (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1549-1553.

Hirota, Y., J. Mordoh and F.Jacob (1970) *J. Mol. Biol* **53**, 369-387.

Hobom, G. (1981) *Curr. Top. Microbiol. Immunol.* **94/95**, 93-142.

Hobom, G., R. Gromchedl, M. Lusky, G. Scherer, E. Schwarz and H. Kossel (1979) *Cold Spring Harbour Sym. Quant. Biol.* **43**, 165-178.

Hobom, G. and M. Lusky (1980) *ICN - UCLA Symp. Mol. Cell. Biol.* **19**, 231-235.

Hocking, S.M. (1977) Ph.D. Thesis, University of Adelaide.

Hocking, S.M. and J.B. Egan (1982a) *Mol. Gen. Genet.* **187**, 1068-1071.

Hocking, S.M. and J.B. Egan (1982b) *J. Virol.* **44**, 1056-1067.

Hocking, S.M. and J.B. Egan (1982c) *Mol. Gen. Genet.* **187**, 87-95.

Hocking, S.M. and J.B. Egan (1982d) *Mol. Gen. Genet.* **187**, 174-176.

Hooper, I. and J.B. Egan (1981) *J. Virol* **40**, 599-601.

Horii, T., T. Ogawa and H. Ogawa (1981) *Cell* **23**, 689-697.

Hooper, I. (1979) Ph.D. Thesis, University of Adelaide.

Hoyt, M.A., D.M. Knight, A. Das, H.I. Miller and H. Echols (1982) *Cell* **31**, 565-573.

Inman, R.B. and G. Bertani (1969) *J. Mol. Biol.* **44**, 533-550.

Inman, R.B. (1981) *Proceedings of the 3rd Cleveland Symposium on Macromolecules* (ed. A.G. Walton) p. 155, Elsevier, Amsterdam.

Jacob, F., S. Brenner and F. Cuzin (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28**, 329-248.

Jacq, A., H. Lother., W. Messer and M. Kohiyama (1980) *ICN - UCLA Symp. Mol. Cell. Biol.* **19**, 189-197.

Kaguni, J.M., R.S. Fuller and A. Kornberg (1982) *Nature (London)* **296**, 623-627.

Kastelein, R.A., E. Remaut, W. Fiers and J. van Duin (1982) *Nature (London)* **295**, 35-40.

Kastelein, R.A. and J. van Duin (1982) *"Interaction of translation and transcriptional controls in the regulation of gene expression"*, Ed. M. Grunberg-Mango and B. Safer, pp. 221-240, Elsevier Amsterdam.

Kimura, M., Y. Takahashi and T. Nagata (1980) *J. Bacteriol.* **144**, 649-655.

Kingsbury, D.T. and D.R. Helinski (1973) *J. Bacteriol.* **114**, 1116-1124.

Klein, A., W. Reiser and A. Anderl (1980) *ICN - UCLA Symp. Mol. Cell. Biol.* **19**, 619-628.

Kobori, J.A. and A. Kornberg (1982a) *J. Biol Chem.* **257**, 13757-13762.

Kobori, J.A. and A. Kornberg (1982b) *J. Biol Chem.* **257**, 13763-13769.

Kogoma, T. (1978) *J. Mol. Biol.* **121**, 55-69.

Kogoma, T. and K. von Meyenburg (1983) *The EMBO journal* **2**, 443-468.

Kogoma, T., T.A. Torrey, M.L. Subia and G.G. Pickett (1981) *ICN - UCLA Symp. Mol. Cell. Biol.* **22**, 361-374.

Kornberg, A. (1980) *"DNA Replication"* W.H. Freeman and Co., San Francisco.

Kornberg, A. (1982) *"1982 Supplement to DNA Replication"*, W.H. Freeman and Co., San Francisco.

Kroger, M. and G. Hobom (1982) *Gene* **20**, 25-38.

Langveld, S.A., A.D.M. van Mansfield, P.D. Baas, H.S. Jansz, G.A. van Arkel and P.J. Weisbeek (1978) *Nature (London)* **271**, 417-420.

Lindahl, G. (1970) *Virology*, **42**, 522-533.

Lindahl, G. (1974) *Mol. Gen. Genet.* **128**, 249-260.

Lindahl, L., M. Yamamoto, M. Nomura, J.B. Kirshbaum, B. Allet and J.-D. Rochaix (1977) *J. Mol. Biol.* **109**, 23-47.

Lindqvist, B.H. (1971) *Mol. Gen. Genet.* **110**, 179-196.

Little, J.W., D.W. Mount, Yaniseh-perroner (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4199-4203.

Ljungquist, E. (1973) *Virology* **52**, 120-129.

Mandel, M. and A. Berg (1968) *Proc. Natl. Acad. Sci. U.S.A.* **60**, 265-268.

Maniatis, T., E.F. Fritsch and J. Sambrook (1982) *"Molecular Cloning, A Laboratory manual"* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Masters, M. and P. Broda (1971) *Nature new Biol.* **232**, 137-140.

Masamune, Y. (1976) *Mol. Gen. Genet.* **149**, 335-345.

Messing, J. and J. Vieira (1982) *Gene* **19**, 269-276.

Messing, J., R. Crea and P.H. Seeburg (1981) *Nucl. Acids Res.* **9**, 309-321.

Messer, W., L. Dankworth, R. Tippe-Schindler, J.E. Womack and G. Zahn (1975) *"DNA Synthesis and its regulation" ICN - UCLA Symp. Mol. Cell. Biol.* **3**, 602-617.

Messer, W., B. Heimann, M. Meijer and S. Hall (1980) *ICN - UCLA Symp. Mol. Cell. Biol.* **19**, 161-169.

Meyer, T.F., K. Geider, C. Kurz and H. Schaller (1979) *Nature (London)* **278**, 365-367.

Miki, T., Y. Ebina, F. Kishi and A. Nakazawa (1981) *Nucl. Acids Res.* **9**, 529-543.

Miller, J.H. (1972) *"Experiments in Molecular Genetics"* Cold Sping Harbor Laboratory, Cold Spring Harbor, New York.

Mitra, S. and D.R. Stallions (1976) *Eur. J. Biochem.* **67**, 37-45.

Murakami, A., H. Inokuchi, Y. Hirota, H. Ozeki and H. Yamagishi (1980) *Mol. Gen. Genet.* **180**, 235-247.

Murray, K. and W.E. Murray (1973) *Nature (London)* **243**, 134-139.

Nogare, J.D. (1980) B.Sc.(Hons) Thesis, University of Adelaide.

Normark, S., S. Bergstrom, T. Edlund, T. Grundstrom, B. Jaurin, F.P. Lindberg and O. Olsson (1983) *Ann. Rev. Genet.* **17**, 499-525.

Ogawa, T. and J. Tomizawa (1968) *J. Mol. Biol.* **38**, 217-225.

Ohmori, H., M. Kimura, T. Nagata and Y. Sakakibara (1984) *Gene* **28**, 159-170.

Oka, A., K. Sugimoto, M. Takanami and Y. Hirota (1980) *Mol. Gen. Genet.* **178**, 9-20.

Oppenheium, D. and C. Yanofsky (1980) *Genetics* **95**, 785-795.

Prescott, D.M. and P.L. Kuempel (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2842-2845.

Pritchard, R.H. (1978) *"DNA Synthesis"* (ed. I. Molineux and M. Kohiyama) Plenum Press, New York and London p 101-111.

Reeve, J.N. (1979) *Methods Enzymol.* **68**, 493-503.

Reeve, J.N. and J.E. Shaw (1979) *Mol. Gen. Genet.* **172**, 271-279

Saint, R.B. and J.B. Egan (1979) *Mol. Gen. Genet.* **171**, 79-89.

Sancar, A., C. Stachelek, W. Konigsberg and W.D. Rupp (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2611-261.

Sancar, G.B., A. Sancar, J.W. Little and W.D. Rupp (1982) *Cell* **28**, 523-530.

Sanger, F., G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, J.C. Fiddes, C.A. Hutchison III, P.M. Slocombe and M. Smith (1977) *Nature (London)* **265**, 687-705.

Sanger, F., A.R. Coulson, G.F. Hong, D.F. Hill and G.B. Petersen (1982) *J. Mol. Biol.* **162**, 729-773.

Sanger, F., A.R. Coulson, B.G. Barrell, A.J.H. Smith and B.A. Roe (1980) *J. Mol. Biol.* **143**, 161-178.

Schnos, M. and R.B. Inman (1982) *J. Mol. Biol.* **159**, 457-466.

Schnos, M. and R.B. Inman (1970) *J. Mol. Biol.* **51**, 61-73.

Schnos, M. and R.B. Inman (1971) *J. Mol. Biol.* **55**, 31-38.

Schnos, M., K.J. Denniston, F.R. Blattner and R.B. Inman (1982) *J. Mol. Biol.* **159**, 441-445.

Schubach, W.H., J.D. Whitmer and C.I. Davern (1973) *J. Mol. Biol.* **74**, 205-221.

Schildkraut, I., C.D.B. Banner, C.S. Rhodes and S. Parekh (1984) *Gene* **27**, 327-239.

Schumperli, D., K. McKenney, D.A. Sobieski and M. Rosenberg (1982) *Cell* **30**, 865-871.

Scott, J.F. and A. Kornberg (1978) *J. Biol. Chem.* **253**, 3292-3297.

Scott, J.R. (1984) *Microbiol. Rev.* **48**, 1-23.

Shine, J. and L. Dalgarno (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342-1346.

Siebenlist, U., R.B. Simpson and W. Gilbert (1980) *Cell* **20**, 269-281.

Skalka, A. (1977) *Curr. Top. Microbiol. Immunol.* **78**, 202-237.

Skalka, A., M. Poonian and P. Bartl (1972) *J. Mol. Biol.* **64**, 541.

Skalka, A. and P. Hanson (1972) *J. Virol.* **9**, 583-593.

Soll, L. (1980) *Mol. Gen. Genet.* **178**, 391-396.

Stormo G.D., T.D. Schneider and L.M. Gold (1982a) *Nucl. Acids Res.* **10**, 2971-2996.

Stormo G.D., T.D. Schneider, L.M. Gold and A. Ehrenfeucht (1982b) *Nucl. Acids Res.* **10**, 2997-3011.

Sutcliffe, J.G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77-90.

Takahashi, S. (1975) *J. Mol. Biol.* **94**, 385-396.

Takahashi, S., C. Hours, A. Chu and D.T. Denhardt (1979) *Can. J. Biochem.* **57**, 858-866.

Tessman, I., J.S. Fassler and D.C. Bennett (1982) *J. Bacteriol.* **151**, 1637-1640.

Tomizawa, J. (1971) *Bacteriophage Lambda* (ed. A.O. Hershey) p. 549. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Tomizawa, J. and T. Ogawa (1969) *Cold Spring Harbor Symp. Quant. Biol.* **33**, 533-551.

Torrey, T.A. and T. Kogoma (1982) *Mol. Gen. Genet.* **187**, 225-230.

Tsurimoto, T. and K. Matsubara (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 681-692.

Ts'O, P.O.P., G.K. Helmkamp and C. Sander (1962) *Proc Natl. Acad. Sci. U.S.A.* **48**, 686-698.

Valenzuela, M.S. (1975) *Biochem. Biophys. Res. Commun.* **65**, 1221-1228.

von Meyenburg, K., F.G. Hanse, E. Riise, H.E.W. Bergmans, M. Meijer and W. Messer (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 121-128.

Walker, G.C. (1984) *Microbiol. Rev.* **48**, 60-93.

Wang, J.C. (1967) *J. Mol. Biol.* **28**, 403-411.

Wechsler, J.A. (1975) *J. Bacteriol.* **121**, 594-599.

Woods, W.H. (1972) Ph.D. Thesis, University of Adelaide.

Woods, W.H. and J.B. Egan (1974) *J. Virol.* **14**, 1349-1356.

Yamamoto, T., T. Tamura, T. Yokota and T. Takano (1982) *Mol. Gen. Genet.* **188**, 356-359.

Younghusband, H.B., J.B. Egan, and R.B. Inman (1975) *Mol. Gen. Genet.* **140**, 101-110.

Yamaki, H., R. Balakrishnan, W. Hendrickson and M. Schaechter (1980) *ICN - UCLA Symp. Mol. Cell. Biol.* **19**, 199.

Yanofsky, C. (1981) *Nature (London)* **289**, 751-758.

Younghusband, H.B. and R.B. Inman (1974) *Virology* **62**, 530-538.

Yuasa, S. and Y. Sakakibara (1980) *Mol. Gen. Genet.* **180**, 267-273.

Zyskind, J., L.T. Deen and D.W. Smith (1977) *J. Bacteriol.* **129**, 1466-1475.