THE EARLY CONTROL REGION OF TEMPERATE COLIPHAGE 186: SEQUENCE AND TRANSCRIPTION STUDIES

A Thesis submitted for

the degree of

Doctor of Philosophy

at the

University of Adelaide

by

Bill Kalionis, B.Sc.(Hons.)

Department of Biochemistry

Adelaide, 1985

FOR MY PARENTS

ATHANASIOS AND ALEXANDRA KALIONIS

M13 Sequencer's Lament

To the tune of "Send in the Clowns"

So this is it,
A few bases to go,
I've tried and I've tried but the technique's so slow.
I've poured my gels,
I've run quite a few,
They're full of bubbles, they leaked, but why I never knew.
But where are the clones?
I've got to have clones,
The end is so near.

Is my broth rich?
Does it look clear?
Contamination is something I always fear.
Are my plaques blue?
They shouldn't be,
No DNA left, I'm down on my knees.
So give me some clones,
I've got to have clones,
The end is so near.

I've had bad preps,
There've been quite a few,
I've tried B.D.H. PEG, fresh buffers, but nothing would do.
And though they say,
Solutions will keep,
In my hands they last, no more than a week.
So give me some clones,
I've got to have clones,
The end is so near.

I've read my gels,
My eyes are quite sore,
There's still sequence missing, of this I am sure.
But there it is,
Finally done,
I've conquered this fragment and now I have won.
What's this I hear?
A voice from the door.
My supervisor wants 10 kb more!
So give me more clones,
I've got to have clones,
Or I'll be here all year!

Bill Kalionis

TABLE OF CONTENTS

TABLE OF CONTENTS (contd.)

TABLE OF CONTENTS (contd.)

TABLE OF CONTENTS (contd.)

TABLE OF CONTENTS (contd.)

TABLE OF CONTENTS (contd.)

SUMMARY

An essential prerequisite to understanding the regulation of gene expression is to characterise RNA transcripts with respect to their size, gene content, direction of transcription and their precise location on the DNA sequence. This thesis describes DNA sequence and *in vivo* transcription studies of the early control region of the temperate coliphage 186.

The 3.6 kb *Pst*I(65.5%-77.4%) DNA fragment, which contains the early control region, was sequenced and analysed with the aid of computer programs. Evidence was presented that identified genes *D* (a tail gene), *B* (late control), *cI* (maintenance repressor) and *int* (integration). A further three potential genes (*CP69*, *CP75* and *CP76*) were also found on the DNA sequence.

Computer analysis located potential transcription promoters and terminators and allowed the prediction of the following RNA transcripts.

(1) The *p*R rightward transcript of at least 800 b initiating from promoter *p*R, containing the predicted genes *CP75* and *CP76*, and extending past the boundary *Pst*I site at 77.4%. This transcript was predicted to the early lytic transcript.

(2) The 2.2 kb leftward *p*L transcript initiating at promoter *p*L, encoding genes *cI*, *int* and *CP69*, and terminating at the proposed *rho*-independent terminator *t*L. This transcript was predicted to be the lysogenic transcript.

(3) The 290 b rightward *p*B transcript initiating at promoter *p*B, encoding the late control gene *B* and terminating at the proposed *rho*-independent terminator *t*B. The *B* gene was also predicted to be expressed from a larger transcript initiating in the late region.

The position of the three promoters pR, pL and pB was confirmed by locating the 5'-end of the associated RNA transcripts. Northern analysis was used to identify and size *in vivo* transcripts from the early control region. The sizes of the transcripts detected were consistent with those predicted from the analysis of the DNA sequence.

Two sites predicted to regulate RNA transcription were studied.

(1) The cI repressor binding site (operator) was identified by the DNA sequence location of three virulent mutants.

(2) A conserved DNA sequence called the B box was found overlapping the pB promoter and also downstream of three P2 late promoters which are controlled by the B gene-product in a P2.186 hybrid. The B box was postulated to represent a binding site for the B protein.

The following features were predicted to be important in the control of phage 186 gene transcription.

(1) The pL (lysogenic) and pR (early lytic) transcripts involved in the lysis/lysogeny decision show an unusual pattern of transcription; the two transcripts initially converge and overlap in the operator region. The pR promoter was predicted to be under direct cI repressor control, based on the DNA sequence position of virulent mutants.

(2) The *int* gene was encoded on the pL lysogenic transcript and was predicted to be expressed in the lysogen. Genes *int* and cI overlap in manner which suggested the two genes were translationally coupled.

(3) Late control gene B was shown to be transcribed in the lysogen by Northern analysis, but evidence was presented that B gene transcription was subject to repressive autogenous control in the lysogen. It was proposed that after prophage induction, DNA replication provides both the template for B protein activation of late gene transcription and increases template numbers which leads to titration of the finite pool of B protein available in the lysogen, thereby inducing transcription of the autoregulated B gene.

This work has shown that the control of the lysis/lysogeny decision, integration and excision, and late gene expression in phage 186 exhibits new control strategies and carries the promise of new mechanisms to effect control of gene expression.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains material not previously published or written by another person except where due reference is made in the text.

My consent is given to this thesis being made available for photocopying and loan.

B. KALIONIS

ACKNOWLEDGEMENTS

The author wishes to thank the following persons who have been associated at some time with this thesis.

Professor W.H. Elliott for his permission to work in the Biochemistry Department.

My supervisor, Dr. Barry Egan for his excellent advice, gregarious nature, and for sharing his personal philosophy.

The members of LAB 19, past and present, for their friendship, infectious enthusiasm and for providing an exciting environment in which to work.

Helena, Iain and Ian for their criticism and editing of this work.

John for help with the figures and photography.

Sylvia and John for their excellent technical assistance. Mary and Ros for keeping up a never-ending supply of clean glassware.

Jutta Waiblinger for her patience whilst typing this thesis.

A special thanks to Helena, who could spot the difference in my ever-changing moods but never had a cross word.

Finally, I would like to thank my family for their encouragement, love and support.

ABBREVIATIONS

cpm	-	counts per minute
kb	-	kilobase (1,000 base-pairs)
bp	-	base-pair
b	-	base
EtBr	-	ethidium bromide
moi	-	multiplicity of infection
RNA	-	ribonucleic acid
Tris	-	Tris(hydroxymethyl)aminomethane
UV	-	Ultraviolet light
A_{600}	-	absorbance at 600 nm
uCi	-	microcurie
kPa	-	kilopascal
ug	-	microgram
ul	-	microlitre
uM	-	micromolar
Kd	-	kilodalton
um	-	micrometre

Other abbreviations are described in Section 2.

SECTION 1

INTRODUCTION

1. INTRODUCTION

1.1 PURPOSE

The molecular biology of temperate coliphage lambda (λ) and particularly the analysis of transcriptional control mechanisms has played a pivotal role in establishing many concepts essential to our understanding of how gene expression is regulated. Principles of molecular action formulated from such studies have been important not only in prokaryotes, but are increasingly finding relevance to eukaryotic systems.

These studies were greatly facilitated by the development of techniques for rapid sequence analysis of DNA and the identification, mapping and sequencing of RNA transcripts. The purpose of the work presented in this thesis is to extend this type of analysis to the temperate coliphage 186, which exhibits a distinctly different gene organisation to that of phage λ and therefore has the potential for revealing new mechanisms by which genes can be controlled.

1.2 TRANSCRIPTIONAL CONTROLS OF PHAGE LAMBDA

1.2.1 The Temperate Response in Phage λ

Upon infection of a sensitive cell, phage λ can follow the lytic pathway where its DNA is replicated, morphogenetic proteins are synthesised and the mature phage particle is assembled. Ultimately the cell is lysed with concommitant release of progeny phage. Alternatively, the phage can enter the lysogenic pathway, where lytic functions are repressed, the phage DNA inserts into the bacterial genome and is propagated as an integral part of the bacterial DNA. Since phage λ can exist in either of these two mutually exclusive states of gene expression, the lysis/lysogeny decision can be viewed as a simple paradigm for cell differentiation.

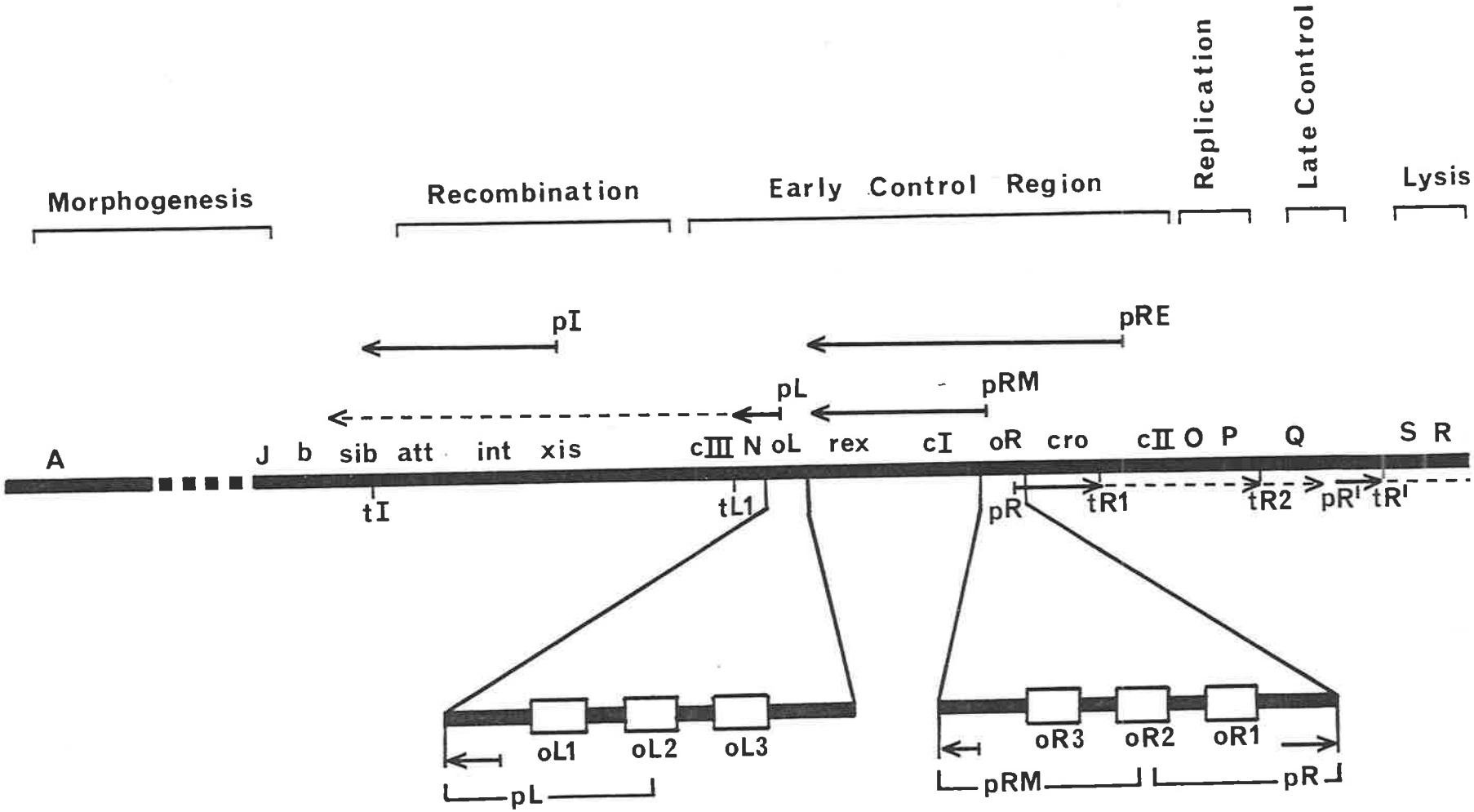
Immediately following injection, λ DNA circularises and transcription by *E. coli* RNA polymerase occurs at promoters *p*L, *p*R and *p*R' (Fig. 1.1). Transcripts from these promoters terminate

Fig. 1.1. Genetic map of phage λ and transcripts in the early region

Identified on the genetic map are representative genes discussed in the text. The clustering of genes which are related by function is indicated above the map. The 40 kb region containing the morphogenetic genes A-J is not shown in full.

Transcription of λ early genes, upon infection, occurs from promoters pL, pR and pR' and terminates at tL1, tR2 and tR' respectively. Some transcription continues to tR2. Transcription of cI (repressor) and rex can be initiated from pRM (promoter for Repressor Maintenance) or pRE (promoter for Repressor Establishment). The cII and cIII proteins stimulate transcription at promoters pI and pRE. Transcripts antiterminated by the action of N (pL, pR transcripts) and Q (pR' transcript) are shown by a dashed line.

The operator regions are expanded to show the spatial relationship of the cI repressor binding sites (at oL and oR) to promoters pL, pRM and pR. Cro repressor also binds at oR and oL to block transcription from these promoters.


Morphogenesis
Recombination
Early Control Region
Replication
Late Control
Lysis
pI
pRE
pL
pRM
A
J b sib att int xis
cIII N oL rex cI oR cro cII O P Q S R
tI
tL1
pR
tR1
tR2
pR'
tR'
oL1
oL2
oL3
pL
oR3
oR2
oR1
pRM
pR

at tL1, tR1 and tR' respectively. Repressor protein Cro, and N protein are synthesised from the pR and pL transcripts respectively. Cro is an enforcer of the lytic pathway and binds to operator regions oR and oL. Binding of Cro to oR blocks premature synthesis of the enforcer of lysogenic development, cI, from the pRM promoter.

The N protein functions as an antiterminator of transcription and acts when E. coli RNA polymerase traverses a region of hyphenated dyad symmetry (the nut site) found downstream from each of the promoters pL and pR (Rosenberg et al., 1978). Efficient modification of transcription at the nut site requires N, and the products of a number of E. coli genes (nusA, nusB, rpoB, rho and nusJ) (reviewed by Herskowitz and Hagen, 1980; Ward and Gottesman, 1982; Friedman et al., 1984). A putative recognition site for one of these proteins (NusA) has been identified in the region immediately promoter-proximal to the nut site and this same site (called box A) has been found near transcription termination sequences in E. coli operons where NusA is involved in transcription termination (Friedman and Olson, 1983).

Antitermination of transcription from pL and pR, past terminators tL1 and tR1 allows transcription of the delayed early genes cII and cIII (necessary for cI transcription), replication genes O and P, and the late turn-on gene Q. Evidence has been provided that at sufficiently high concentrations, Q protein acts by antitermination of the pR' transcript at terminator tR' (Forbes and Herskowitz, 1982) to stimulate late gene transcription in a manner analogous to the action of N protein. A qut site (analogous to the nut site) has been mapped to the immediate right of the pR' promoter (Somasekhar and Szybalski, 1983).

In summary, after initial transcription from promoters pL, pR and pR', cells contain the necessary proteins to execute either the lytic or lysogenic pathways and the phage is therefore considered to be at the uncommitted growth stage. Transcription from the promoter for cI transcription in the prophage state, requires cI itself for its activation (See Section 1.2.3), however Cro repressor bound to the oR operator ensures transcription from pRM does not occur immediately upon infection. Transcription of cI can occur from a second promoter, pRE, the activity of which is dependent on the action of the cII and cIII proteins. A second promoter, pI, is also stimulated by the action of cII and cIII.

1.2.2 The cII Protein and its Role in the Lysis/Lysogeny Decision

The level of cII protein in the cell is believed to be the key determinant in the choice between the lytic and lysogenic pathways, through its action in stimulating transcription from pRE and pI, and through a poorly understood mechanism whereby cII protein synthesis causes a delay in lytic growth (McMacken et al., 1970).

Promoters pRE and pI have been sequenced and although neither show extensive homology to the prokaryotic consensus promoter, they do share hołology in the -35 region (the -10 region shows little homology) (Wulff and Rosenberg, 1983). Analysis of pRE mutants which affect cII binding revealed a clustering of the mutations into two 4 bp direct repeats (TTGC) flanking the -35 hexamer (Ho and Rosenberg, 1982). DNAse protection studies and methylation studies provided evidence that cII selectively interacts with the TTGC repeat (Ho et al., 1983).

The level of cII protein in the cell is determined by the interaction of cII with host proteins HflA and HflB (for extensive reviews see Herskowitz and Hagen, 1980; Wulff and Rosenberg, 1983;

Friedman et al., 1984) which act as antagonists of lysogenisation and are considered to be proteases, or control proteolytic activities that degrade cII. Two further influences on the level of cII in the cell are the λ cIII protein which responds to the multiplicity of infection, and the E. coli catabolite activation system (CAP and cAMP) which responds to the nutritional status of the cell. cIII and CAP-cAMP are considered to be negative regulators of the Hfl proteins (or the proteolytic activities under Hfl control). High multiplicities of infection and starvation of the cells prior to infection increase the levels of cIII and CAP-cAMP respectively and consequently increase the level of cII in the cell through their inhibitory effects on the action of Hfl. A further influence on the levels of cII in the cell are the proteins HimA and Hip-HimD, but, unlike Hfl and cII, these proteins do not affect cII stability and their action is poorly understood.

In summary, various environmental factors such as host physiology and multiplicity of infection influence the level of cII protein in the cell. Herskowitz and Hagen (1980), and Echols (1980) argue that the level of cII is the crucial determinant of whether the committment is made to the lytic or lysogenic pathway since it co-ordinates two genetic switches to produce a unified lytic or lysogenic response.

Competition of repressors cI and Cro to repress each others transcription at the operator site oR (the cI/Cro switch) leads to two stable and mutually exclusive conditions (lytic functions off or on) and thereby constitutes a genetic switch. The other genetic switch, comprised of the int and xis genes and two DNA sites (att and sib), determines whether the phage chromosome integrates into the host genome or not (Integration/Excision switch). Co-ordination of these two switches by cII ensures that both switches are thrown

in the same direction i.e. lytic functions are on when the phage is unintegrated and lysogenic functions are on when the phage is integrated. Any other combination of these switches would lead to loss of the prophage or cell death. The two switches and the manner in which they are co-ordinated by cII will be discussed.

1.2.3 The cI/Cro Switch

Phage λ contains two operator regions (oL and oR) where control of early gene transcription occurs. The cI/Cro switch, which determines whether lytic functions are on or off, acts primarily at the oR operator (Ptashne et al., 1980; Herskowitz and Hagen, 1980; Meyer and Ptashne, 1980; Dambly-Chaudiere et al., 1983; Shea and Ackers, 1985). Operator oR is located between two genes, cI and cro, which encode repressor proteins that play an essential role in lysogenic and lytic development respectively. The operator region is tripartite, consisting of three closely-spaced, non-overlapping 17 bp DNA sequences (named oR3,oR2,oR1) which show inverted repeat symmetry (Maniatis et al., 1975; Ptashne et al., 1976) and each operator site is able to bind both cI and Cro proteins.

Dimeric forms of the cI and Cro repressor proteins exist in equilibrium with monomeric forms of the two proteins, however only the dimeric forms of the two repressors bind to the operator sites (Ptashne, 1971). The amino-terminal region of the cI repressor and the Cro protein have been analysed by X-ray crystallography and their structures determined (Pabo and Lewis, 1982; Anderson et al., 1981). Although the two proteins recognise different DNA sequences in the operator region, they show structural similarities in the proposed DNA-binding region of the two proteins, which is believed to be a pair of α helices named the α2-α3 helices (Steitz et al., 1982; Matthews et al., 1982; Ohlendorf et al., 1982, See Section 8.1 for a more detailed discussion).

The two repressors have reciprocal affinities for each operator site and this property is of crucial importance to the functioning of the cI/Cro switch (Johnson et al., 1981). Cro binds first to oR3 and at higher concentrations to oR2 and oR1 whilst cI fills oR1 and oR2 together and oR3 is filled only at higher concentrations (Johnson et al., 1979). The binding of cI at the oR operator is influenced by a co-operative interaction that occurs between adjacently bound cI molecules, such that cI repressor binding at oR1 co-operatively facilitates the binding of cI to oR2 (but not oR3).

Two divergent promoters are found in the oR operator region; the pR promoter for the primary rightward pR transcript which encodes the cro gene and the pRM promoter, for the maintenance repressor gene cI, which initiates leftward transcription. The pR and pRM promoters overlap operators oR1 and oR3 respectively (Fig. 1.1). This arrangement of a tripartite operator region allows the two repressor proteins (cI and Cro), which have reciprocal affinities for the three operator sites, to complete and block transcription of each others gene.

1.2.4 The Integration/Excision Switch

Integration and excision of the λ genome during lysogeny and induction results from a reciprocal site-specific recombination event between the phage attP and bacterial attB sites (Campbell, 1962). The attachment sites share a common 15 bp core region (Landy and Ross, 1977) and the crossover event occurs within this common core region (Mizuuchi et al., 1981).

Pivotal to the control of the Integration/Excision switch is the requirement for Int (integrase) alone among the phage-encoded proteins (Gottesman and Yarmolinsky, 1968) for integration whereas both Int and Xis (excisionase) are required for excision (Guarneros and Echols, 1970). Int is a topoisomerase and it executes the breakage

and rejoining reactions of strand exchange (Craig and Nash, 1983).

Genes int and xis are the last two genes in the left operon of λ and the xis gene transcriptionally precedes the int gene. The coding sequences of the two genes overlap (Hoess et al., 1980) and the promoter pI is located at the beginning of the xis gene such that the pI transcript lacks the first few codons of Xis. Consequently Int but not Xis is synthesised from the pI transcript. Transcription from the pI promoter terminates 277 nucleotides beyond the int gene at the tI terminator (Schmeissner et al., 1984a).

Both int and xis genes are transcribed from the pL promoter but promoters pI and pL are positively regulated by different effectors; pI is activated by the cII protein (Schmeissner et al., 1981, Ho et al., 1983) and results in int gene transcription, whilst the N gene-product (encoded on the pL transcript, See Section 1.2.1) antiterminates transcripts initiating at pL, allowing read-through into the xis and int genes. The pL promoter is also under negative control of λ repressor.

The differential requirement of Int for integration (Int alone) and excision (Int and Xis) can be accomplished by selective transcription from pI and pL respectively.

1.2.5 Co-ordination of the Two Switches

As discussed previously (Section 1.2.2) the level of the cII protein is believed to be the crucial determinant in the committment to execute the lysis or lysogeny pathways.

High levels of cII favor the lysogenic response, firstly by activating cI transcription from pRE and int transcription from pI and secondly by delaying lytic growth. The synthesis of cI leads to displacement of Cro from the operator sites. Binding of cI to oR1 and oR2 turns off transcription of cII, O and P from pR, and binding to oL turns off N and cIII transcription. Co-operative

binding of cI to oR1 and oR2 leads to a rapid increase in cI repressor concentration and provides favourable protein-protein contact with RNA polymerase thereby positively stimulating pRM transcription (Sauer et al., 1979; Ptashne et al., 1980, Meyer and Ptashne, 1980). Synthesis of Int protein from the pI promoter (stimulated by cII) allows the phage to integrate into the bacterial genome and complete the lysogenic pathway.

Low levels of cII favor the lytic pathway since transcription from pR predominates over that from pRE and allows Cro levels to increase in preference to cI, thereby allowing transcription of N, O, P and Q. As the level of Cro increases early gene transcription is turned off by binding of Cro to oR2 and oR1 to block pR transcription and by binding to oL2 and oL1 to block pL transcription.

1.2.6 Retroregulation at the sib Site

During lytic development both Int and Xis can be expressed from the pL transcript even though neither is required for lytic development. The synthesis of Int under these conditions is potentially deleterious should integration occur during a productive infection. The mechanism of downstream retroregulation (Miller et al., 1981; Guarneros et al., 1982) of the sib site prevents the expression of int during lytic development.

Deletion analysis of the sib site (Court et al., 1983) and the DNA sequence analysis of sib mutants (Guarneros et al., 1982) have allowed the precise location of the sib site to be determined. The sib site is a palindromic sequence located to the left of attP and distal to the int gene. Transcripts initiating at pL are prevented from termination at terminator tL1 (the first terminator encountered by the pL transcript) and terminator tI, through the action of the N gene-product which modifies RNA polymerase complexes that initiate at pL (See Section 1.2.1). The pL transcript extends into

the b region beyond tI, and this extended transcript forms an RNAse III processing structure at the sib site (Schmeissner et al., 1984b). RNAse III processing of the pL transcript at sib is proposed to lead to the decay of the int mRNA through the action of a 3' to 5' exoribonuclease (Schmeissner et al., 1984b). Retroregulation at the sib site therefore prevents int expression from pL during the lytic pathway of development.

The pI transcript does not contain the site necessary for the action of the N gene-product (the nut site) and therefore terminates at tI, before sib. Consequently, this transcript expresses int efficiently. Retroregulation thus ensures that, upon infection, int is expressed efficiently only from the pI transcript.

Processing of the sib site affects int synthesis from the pL transcript but not the expression of xis which is located upstream of int (Schindler and Echols, 1981), and the synthesis of excisionase in an infected cell before commitment would be potentially disadvantageous should the decision be made to follow the lysogenic pathway. Weisberg and Gottesman (1971) provided evidence that Xis decays rapidly, whereas Int is stable. Metabolic instability of Xis and retroregulation are regulatory mechanisms used by λ to control the insertion and excision pathways.

1.2.7 Maintenance of the Prophage State

In addition to repression of transcription at pR (and pL) the cI repressor must stimulate transcription of its own gene from the pRM promoter, to maintain the lysogenic state. This is achieved by the repressor dimer bound at oR2, which in turn is stabilised by repressor dimer bound to oR1 (See Section 1.2.3). The concentration of cI must be maintained at a steady state to avoid random fluctuations to inducing levels, and is achieved by the synthesis of a large amount of cI mRNA from the fully stimulated pRM promoter, but poor

translation of the resulting transcript (Ptashne *et al.*, 1976).

1.2.8 Induction from the Prophage State

Lysogenic induction requires a *recA*$^+$ host and whilst the RecA protein is the principal enzyme that catalyses general recombination, it also participates in the activation of a bank of genes (the SOS genes) which are involved in DNA repair (Little and Mount, 1982). UV-irradiation of the host activates the RecA protein which leads to proteolytic cleavage of the repressor of *recA* transcription, LexA (Little *et al.*, 1980), and of the λ repressor monomer (Roberts *et al.*, 1978) in the hinge region between its two structural domains (Sauer *et al.*, 1979). In the case of the λ repressor, the free amino-terminal repressor fragments produced by the cleavage are incapable of efficient dimerisation, and tight binding to the operator does not occur (Pabo *et al.*, 1979). The *p*R promoter is the first phage promoter to be derepressed since repressor dimers bind less tightly to the operator sites at *o*R than those at *o*L (Johnson *et al.*, 1981). Cro protein synthesised from *p*R binds tightly to *o*R3 to prevent further *p*RM-dependent *c*I transcription and this event is most likely to result in committment to the lytic pathway of development.

Excision from the prophage state requires both Int and Xis (See Section 1.2.4) which are synthesised from the *p*L transcript on induction. Since *sib* is located to the left of *att*, the reciprocal site-specific recombination event that occurs on integration, places *sib* at the opposite end of the chromosome to *int*. Retroregulation at the *sib* site cannot occur upon prophage induction and both Int and Xis are synthesised from the *p*L transcript, allowing excision to proceed.

1.2.9 Transcriptional Controls of the Lambdoid Phages

The lambdoid family of phages are defined by their ability to exchange genetic information with phage λ via homologous recombination and includes phages φ80, 21, 434, 424 and P22. Members of this family have a similar genetic organisation, reflected in the gene order and the relative position of the genes with similar functions along the genome. Regulatory organisation of the lambdoid family is important in allowing homologous recombination to generate functional recombinants with different specificities.

Transcriptional control mechanisms of these phages have common features:

(1) a positive regulator (λ N-like or Q-like) which acts to antiterminate transcripts of operons.

(2) early lytic functions form two divergently transcribed operons, flanked by two operator regions, which are negatively controlled by a repressor encoded among the genes between the two operons.

(3) a λ cro-like function to antagonise repressor transcription and a λ cII-like function to foster transcription of the maintenance repressor.

(4) Mascarenhas et al. (1983) provided evidence that phage 434, like λ shows the sib control mechanism of int gene expression and suggest that this control feature may also be common in the lambdoid family.

(5) important features of transcriptional control found throughout the lambdoid family include the proximity of regulatory genes to their site of action and the regulation of long operons by sites near the initiating points of transcription. These features allow regulatory specificity to be altered during the formation of functional recombinants by homologous recombination of relatively small regions.

Further study of the mechanisms that control gene expression in the lambdoid family will increasingly reveal differences in regulatory specificity. However, an appreciation of the basic parameters required for any temperate phage to decide between lysis or lysogeny, execute either pathway, and a more complete understanding of the host-virus relationship, must come from a study of temperate phages that show a significantly different gene organisation to that of the lambdoid family. Studies of such phages have the potential of revealing new mechanisms for the control of gene expression.

1.3 PHAGE 186

The P2-related family, which includes phages P2, 186, P4, 299, 18 and W, form a second group of temperate phages (Bertani and Bertani, 1971) which are distinct from the lambdoid family. Phage 186, like other members of the P2-related family, shows a different arrangement of genes involved in the establishment and maintenance of lysogeny, and induction of the lysogen, to that found in lambda. Additionally, phage 186 has several properties which distinguish it from the prototype phage of this family, P2 (Bertani and Bertani, 1971), which will be detailed in later discussion.

Bacteriophage 186 has a double-stranded, non-permutated DNA genome of approximately 30 kb (Wang and Schwartz, 1967; Younghusband et al., 1975; Chattoraj et al., 1973) and possesses complementary cohesive ends (cos) (Baldwin et al., 1966; Wang et al., 1973).

A linear genetic map (Hocking and Egan, 1982a) and a physical map (Saint and Egan, 1979) were constructed and the gene content of cloned DNA restriction fragments was determined (Finnegan and Egan, 1979), allowing the physical and genetic maps to be orientated with respect to each other. Phage 186 DNA strands, separated by CsCl buoyant density gradient centrifugation, were orientated with

respect to the physical and genetic maps by Kalionis and Egan (1981).

1.3.1 Genetic Organisation of Phage 186

Genes with related functions are clustered on the 186 genetic map (Fig. 1.1(a)), with at least 22 essential genes for lytic growth lying in the left 0% to 65% of the genome (Hocking and Egan, 1982b). Phage tail genes are represented by genes N to D and head genes by W to Q. Gene P is the lysis gene, whilst the function of gene O is not known. The left 0% to 65% of the genome contains four polarity groups, as defined by the polar effects of various amber mutations (Hocking and Egan, 1982b).

The lysis/lysogeny decision, replication of phage DNA and late gene transcription are regulated by genes located in the early control region of the genome (65% to 100%) and will be discussed in further detail below.

1.3.1(a) Region att-int-cI

The position of the int (integration) and cI (repressor) gene was indicated by the mapping of two insertion mutants, ins3 at 70.3% and ins1 at 73.5%, (Younghusband et al., 1975; Bradley et al., 1975) which show the Int^- and cI^- phenotypes respectively.

Recombinational mapping placed the prophage attachment site, att, between genes B and cI (Hocking and Egan, 1982a), but the position of att relative to the int gene was unknown. The att site was placed at 67.9%, since this represents the left boundary of the deletion mutant Δ1, a mutant formed by mutagenesis during prophage induction (Dharmarajah, 1975). Deletion mutant Δ1 extends from 67.9% to 74.9% (Finnegan and Egan, 1981), shows the Int^-cI^- non-virulent phenotype and approximates the att-int-cI region. This region of approximately 2.1 kb was not expected to encode any other genes, and certainly no essential genes, since the Δ1 mutant is viable.

In contrast, the att-int-cI region of phage λ is approximately 8 kb long and encodes many genes. These genes include the primary control gene N, genes involved in the establishment and maintenance of repression, cIII and cI respectively, and the excision gene xis.

1.3.1(b) Region cI-A-cos

The region between the cI gene and the cohesive ends (cos) represents over one quarter of the 186 genome (about position 74% to 100%); only two genes, A, and dhr, have been located in this area by genetic means (Hocking and Egan, 1982a; Richardson and Egan, 1985).

The cI repressor binding site (O) was placed to the right of the cI gene (at approximately 75%) since the Δ1 deletion mutant gives the cI^- phenotype but is non-virulent, and transcription which initiates to the left of the PstI site at 77.4% was under cI repressor control (Finnegan and Egan, 1981).

Bacterial DNA synthesis is depressed immediately after heat induction of a 186 prophage (Hocking and Egan, 1982c), and a 186 gene dhr (previously named dho) was proposed to be responsible for this effect. This gene has been identified by mutation, found to be non-essential for phage growth, and mapped between 77.4% to 79.6% of the genome (Richardson and Egan, 1985).

Replication gene A spans the region 83.8% to 92.0% since all known Aam alleles map between these co-ordinates (Finnegan and Egan, 1979) and the origin of 186 DNA replication was mapped at 92.9%$\pm$1.8% (Chattoraj and Inman, 1973).

The analogous region of phage λ contains genes cro and cII, the replication genes O and P, late control gene Q, and the lysis genes S and R. Genes cro, cII and cIII, together with various host

factors, are involved in the lysis/lysogeny decision. Although phage 186 has no known equivalent of a λ cro function, amber mutants that fail to establish lysogeny and are able to complement 186 cI mutants, identify a 186 cII gene. This gene is known to map to the right of the 186 att site (Hocking, 1977). Two further complementation groups cIII and cIV have been identified (I. Lamont, personal communication), suggesting that three genes (cII, cIII, cIV) may be involved in the establishment of repression in phage 186.

1.3.1(c) Late Control Gene B

Gene B is required for phage 186 late gene transcription (Finnegan and Egan, 1981) and the two known alleles of the B gene (Bam17, Bam57) map between the tail gene D and the att site (Hocking and Egan, 1982a). A coding potential of 250 amino acids is estimated for the region from the PstI site at 65.5%, (about which all four alleles of the D gene map (Finnegan and Egan, 1979)), and the att site (approximately 68%). No genes other than B and part of the D gene are expected in this region.

Like the Q gene product of λ, the 186 B gene product is a positive regulator of late gene transcription (Finnegan and Egan, 1981), but the two genes differ markedly in their location on their respective genomes and in the mechanism by which they are proposed to act. As described previously (Section 1.2), the Q gene is located at the right end of the λ genome (90.5%) and the Q gene product acts near its site of synthesis by preventing with transcription termination. This gives rise to a single transcript encoding the adjacent lysis genes, S and R, and the late genes. The 186 B gene is located between the tail gene D and the att site (approximately 68%) and, as will be discussed in detail in Section 9, acts at a distance from its site of synthesis and most probably brings

about late gene transcription in 186 by positively controlling several late promoters rather than by antitermination of transcription.

1.4 COMPARATIVE BIOLOGY OF PHAGES 186 AND P2

1.4.1 Genetic Organisation of Phages 186 and P2 (Refer to Fig. 1.2)

Phage 186 shares many properties with P2; the phage virions are very similar (Bertani and Bertani, 1971) and the two similarly-sized phage genomes possess near-identical cohesive ends (Murray et al., 1977; Wang et al., 1973; Murray and Murray, 1973; Pabmanabhan and Wu, 1972).

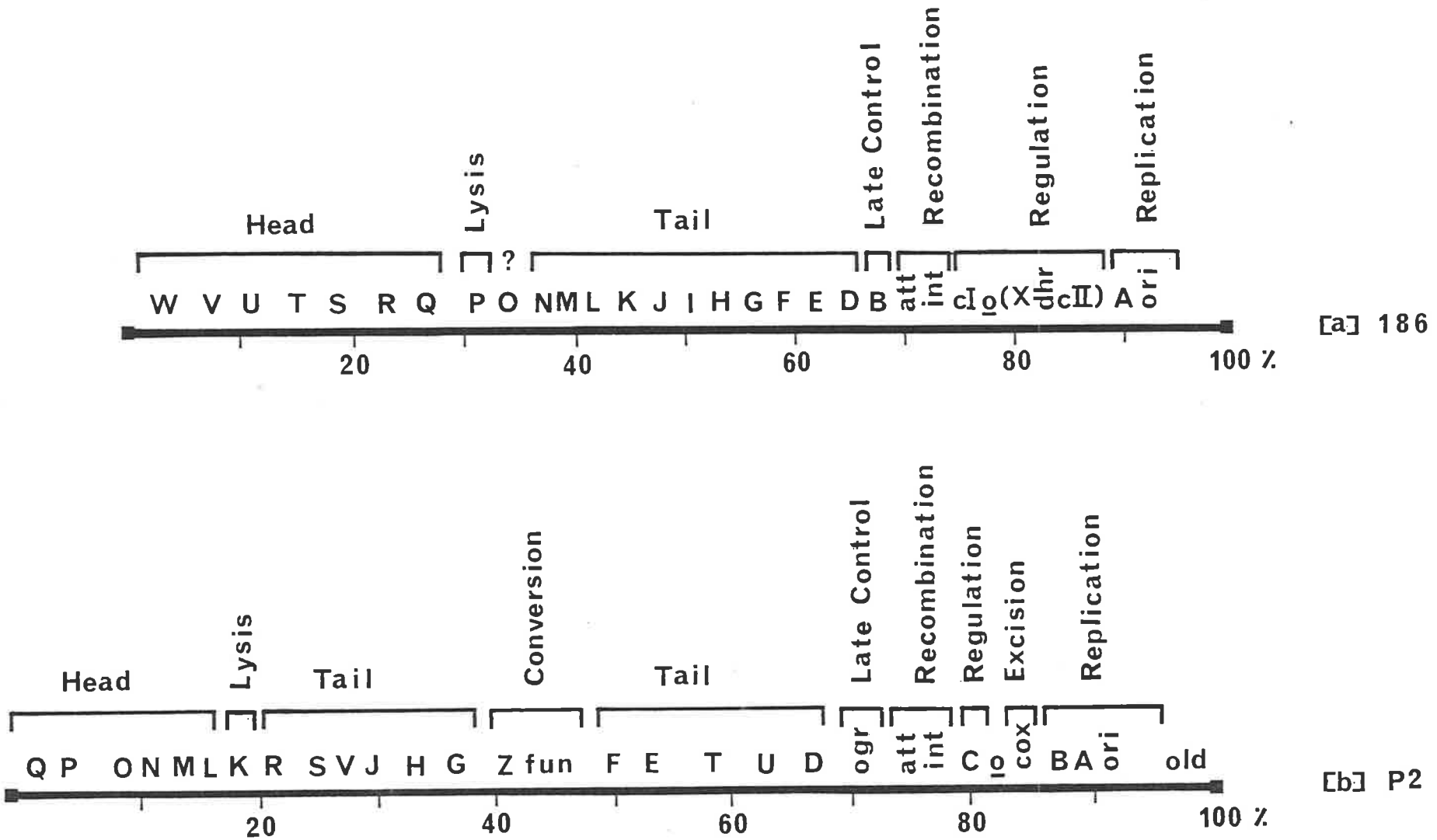
Phage 186 and P2 have a similar arrangement of genes with related functions (Lindahl, 1969; Lindahl, 1974; Hocking and Egan, 1982a); an arrangement which is most important to the work presented in this thesis since it is significantly different to that found in the lambdoid family. The two phages show considerable homology, particularly in the left two thirds of their genomes where structural genes are encoded, as judged by formation of heteroduplexes between the two genomes (Younghusband and Inman, 1974), and the isolation of viable P2.186 hybrids containing P2 late and 186 early genes (Younghusband et al., 1975; Bradley et al., 1975; Hocking and Egan, 1982d). Significant heteroduplex formation does not occur between the early regions (right one third) of the two genomes, where regulatory functions are encoded.

Phages 186 and P2 are temperate phages and so must make a decision between lysis or lysogeny. For the lytic pathway to be executed, some means of repressing lysogenic functions must be available. In phage λ, the enforcer of the lytic pathway is the cro gene-product (Echols et al., 1973), however no cro-like function could be found in P2 (Nilsson and Bertani, 1977) and no known cro-like function has yet been identified in phage 186. P2

Fig. 1.2 Genetic maps of temperate coliphages 186 and P2

(a) Genetic map of phage 186. Functions of the genes were described by Hocking and Egan (1982a). The early control region contains the cI (maintenance repressor), int (integrase) and B (late turn-on) genes. Gene X was predicted by Finnegan and Egan (1981) as being necessary for rightward transcription into the replication gene (A). The cII gene is required for the establishment of lysogeny (Hocking, 1977) and the predicted location of the phage attachment site (att) and operator region (O) is also shown.

(b) Genetic map of phage P2. Functions of the genes were described by Bertani et al. (1982). The early control contains the C (maintenance repressor), int (integrase), cox (excisionase), and ogr (late turn-on) genes. Genes A and B are required for replication. The predicted location of the phage attachment site (att) and operator region (O) is also shown.


Head
Lysis
?
Tail
Late Control
Recombination
Regulation
Replication
W V U T S R Q P O N M L K J I H G F E D B att int cI o (X dhr cII) A ori
20
40
60
80
100 %
[a] 186
Head
Lysis
Tail
Conversion
Tail
Late Control
Recombination
Regulation
Excision
Replication
Q P O N M L K R S V J H G Z fun F E T U D ogr att int C o cox B A ori old
20
40
60
80
100 %
[b] P2

encodes the cox gene (for which phage 186 has no known equivalent) in the analogous position to the λ cro gene and the cox gene-product is required for (inefficient) excision of P2 from the prophage state (Lindahl and Sunshine, 1972).

P2 clear plaque-forming mutants fall into two complementation groups which both appear to be necessary for the maintenance of the P2 lysogen (L.E. Bertani, 1968; Bertani, 1976). Unlike phage 186, a cII-like gene for the establishment of lysogeny has not yet been identified in P2.

After infection of a host cell with phage P2, bacterial DNA synthesis does not appear to be depressed (Hooper 1979), suggesting that a gene analogous to the dhr gene of phage 186 does not exist in phage P2.

DNA replication in P2 is initiated from a single-strand nick introduced by the action of the A gene-product (Chattoraj, 1978; Geisselsoder, 1976). The A gene-product of phage 186 probably functions in a similar manner (A.V. Sivaprasad, 1984). However the B gene-product of phage P2, which is involved in lagging-strand DNA synthesis during replication (Funnell and Inman, 1983), has no known analogous gene in 186.

The constitutive gene old of phage P2 maps at the extreme right end of the P2 genome and is of considerable interest, not only because it causes interference with phage λ growth (Lindahl et al., 1970), but also because it kills recB$^-$ mutants of E.coli after P2 infection and causes increased sensitivity of P2 lysogens to X-ray irradiation (Ghisotti et al., 1983). A 186 prophage does not interfere with phage λ growth, but interference does occur upon mixed infection of 186 and λ (M. Verma, personal communication), suggesting an old-like function in phage 186, which unlike the old gene of P2, appears to be under cI repressor control.

Gene ogr is required for phage P2 late gene transcription (Sunshine and Sauer, 1975) and acts analogously to the 186 B gene. These two genes map in the same relative position on their respective genetic maps (Bertani et al., 1982; Sunshine and Sauer 1975; Hocking and Egan, 1982a) and their expression is controlled by the phage replication (A) genes (B. Sauer et al., 1982; Finnegan and Egan, 1981). The P2 Ogr protein is presumed to function by activating the late promoters (Sunshine and Sauer, 1975; Christie and Calendar, 1983, 1985). These same promoters are presumably activated by the 186 B gene-product as a necessary consequence of the existence of viable P2.186 hybrid phages (Hy 9,10,11-Hocking and Egan, 1982d), where P2 ogr is absent and P2 late gene expression is under 186 B gene control.

1.4.2 Phage 186 and P2 Induction From the Prophage State

The most notable difference between phages 186 and P2 is in their ability to induce from the prophage state. Agents that induce the λ prophage, such as UV-irradiation, nalidixic acid and mitomycin C, will also induce a 186 prophage (Woods and Egan, 1974). These treatments do not result in P2 prophage induction (G. Bertani, 1968). Contributing factors to P2 non-inducibility include the failure of the P2 repressor to be inactivated by UV-irradiation (G. Bertani, 1968) and the failure of the prophage to excise from the host chromosome. The P2 prophage fails to excise even when repression is lifted in a prophage with a temperature-sensitive repressor, when placed at the non-permissive temperature (L.E. Bertani, 1968).

Site-specific integration of phage P2 requires a functional int gene (Choe, 1969) whereas spontaneous prophage excision, a very inefficient process (L.E. Bertani, 1968), requires both int

and the product of the cox gene (Lindahl and Sunshine, 1972). The inability of the P2 prophage to excise was thought to reflect poor expression of the P2 int gene in the prophage state (Bertani, 1970).

A "split-operon" model was proposed by Bertani (1970) to explain how int gene expression could be altered in the prophage state. In this model, the int gene was physically separated from its promoter by the phage attachment site (att), and upon integration of the phage into the host chromosome, the int gene would be split from its promoter, preventing transcription of the int gene. Rightward transcription of the int gene with respect to the genetic map was essential for this model to be valid (since the gene order att-int was known). However, the mapping of the int150 and int239 amber peptide fragments of the P2 Int protein (Ljungquist and Bertani, 1983) and the recently determined DNA sequence of the int239 amber mutant (E. Ljungquist, personal communication), confirmed that the P2 int gene was in fact transcribed in the opposite direction, invalidating the "split-operon" model.

Phage 186 is inducible from the prophage state and int gene expression is required for site-specific integration (Younghusband et al., 1975) and for excision from the prophage state (Bradley et al., 1975) however an excision function analogous to P2 cox or λ xis has not been identified by genetic studies.

1.4.3 Satellite Phage P4 and its Interaction with Phages 186 and P2

P4 is a satellite phage of the P2-related family consisting of an 11 kb linear, non-permutated, double-stranded DNA genome (Inman et al., 1971) and, while the 19 nucleotide long cohesive ends are identical to those of phage P2, the genome of the two phages are essentially heterologous (Lindqvist, 1974).

As a temperate phage, P4 can follow either the lytic or

lysogenic pathway, and a single spontaneous mutation (vir1) allows P4 to be maintained as a plasmid (Shore et al., 1978; Goldstein et al., 1982).

Lytic development of P4 requires all known head, tail and lysis genes of a helper phage such as P2 (Six, 1975) or 186 (B. Sauer et al., 1982). DNA replication and lysogenisation are independent of a helper phage (Lindqvist and Six, 1971). P4 is of considerable interest because it encodes several functions, ε, δ, psu and sid (Geisselsoder et al., 1981; Souza et al., 1977; Sauer et al., 1981; Diana et al., 1978), which act to alter the control of transcription of the helper phage.

The late functions required by P4 can be provided either by a co-infecting P2 phage or by a resident P2 prophage (Six and Klug, 1973). P4 can utilize P2 late genes by derepressing the P2 prophage which results in the expression of P2 late genes under their normal mode of control (Six and Lindquist, 1978; Geisselsoder et al., 1981). Derepression of the P2 prophage is dependent on the P4 ε gene-product and results in in situ unidirectional replication of the P2 prophage helper without its excision from the host genome (Geisselsoder et al., 1981). Additionally, P4 can activate expression of P2 late genes by "transactivation" (Souza et al., 1977), a process whereby P4 causes the transcription of P2 late genes in the absence of functional P2 A and B gene-products (Six, 1975; B. Sauer et al., 1982), which are normally required for P2 late gene expression (Lindahl, 1970; Geisselsoder et al., 1973). The P4 transactivation gene is called δ (Souza et al., 1977).

Phage 186 can also act as a helper phage for P4 during a mixed infection and P4 is able to transactivate 186 late gene expression from a 186 lysogen in the absence of functional 186 A and B gene-products (B. Sauer et al., 1982), which are normally required for 186 late

gene expression (Finnegan and Egan, 1981). The utilisation of late gene functions from a 186 prophage by P4, requires that the 186 prophage be derepressed (e.g. by temperature induction) (B. Sauer et al., 1982), presumably because the P4 ε gene product does not cause 186 prophage derepression. The simplest interpretation of this requirement for 186 prophage derepression is that P4 requires a 186 function under 186 cI repressor control, which is essential for P4 lytic infection to proceed.

1.5 IN VIVO TRANSCRIPTION STUDIES OF PHAGES 186 and P2

The discussion to this point has highlighted the similarities and differences in the overall genetic organisation of phages from the two main temperate phage families (lambdoid and P2-related) and from within the P2-related family (186, P2, P4). The control of early gene transcription in vivo is of primary concern to the work presented in this thesis, and an essential prerequisite is to characterise the in vivo transcripts of the early control region.

Bacteriophages 186 and P2 have been well characterised both genetically and physically, but comparatively little is known about the regulation of early gene transcription, and even less about the in vivo RNA transcripts from the early region.

1.5.1. Phage P2

In vivo transcription studies of the early control region of phage P2 have been limited to RNA/DNA hybridisation experiments that employed separated DNA strands or sheared halves of P2 DNA (Lindqvist and Bøvre, 1972; Geisselsoder et al., 1973), or R-loop analysis (Funnell and Inman, 1982). P2 transcription was shown to be predominantly left to right with respect to the genetic map, with a small but significant amount of transcription in the opposite direction. In the early stages of P2 infection, transcription was predominantly from the right half of the genome (early region)

but as infection proceeded, a transition of transcription to the left half of the genome (late region) occurred. Gene-products of the two known replication genes A and B were necessary for this transition. Funnell and Inman (1982) concluded from R-loop analysis data that the promoter for the early operon in P2 was located at 77%±1.4% on the P2 DNA map. Further characterisation of the P2 *in vivo* RNA transcription pattern from the early control region has not been reported.

1.5.2 Phage 186

Finnegan and Egan (1981) hybridised pulse-labelled RNA, isolated at different times after prophage induction (by temperature increase), to cloned DNA restriction fragments of known gene content (Finnegan and Egan, 1979) and determined the segmental distribution of RNA transcription on the 186 genome.

Transcription from the prophage was from the region 67.9% to 74.9%, which was expected since this region encodes the cI maintenance repressor, but unexpectedly the region 92.0% to 93.8% was also transcriptionally active in the prophage. Transcription from the region 92.0% to 93.8% remained low throughout the period after heat induction of the prophage.

Upon induction, transcription was predominantly from the early region (65.4%-100%), reaching a maximum at 30 minutes. After this time the rate of transcription decreased markedly, suggesting a negative control of early gene transcription. Late gene transcription was detected 20 to 25 minutes after induction, as judged by hybridisation to clones representative of the region 0% to 65.4%, and continued until cell lysis.

Induction of a 186 prophage with a defective replication gene A (Aam) had two distinct effects; late gene transcription was essentially absent after induction of the Aam prophage and

transcriptional activity of the pEC35 region (Fig 1.3), which was assumed to reflect transcription of the late control gene B, decreased dramatically and led to the conclusion that B gene transcription was dependent upon A gene function.

After induction of a Bam prophage, early gene transcription resembled that of the wild type phage for 25 minutes but after this period of time the normal transition to late gene transcription did not occur, instead the rate of transcription from fragments pEC18, pEC17.2 and pEC15 increased by two to five fold (see Fig. 1.3). Hocking and Egan (1982c) investigated the rate of DNA replication after induction of the Bam prophage and found that replication continued to increase at the same rate as in the wild type phage for at least 60 minutes, whereas replication of the wild type phage shows a distinct decrease in the rate of DNA replication after 30 to 35 minutes. From the above results, it was concluded that the B gene controls late gene transcription and either directly or indirectly, plays a role in the turn-off of early gene transcription and DNA replication.

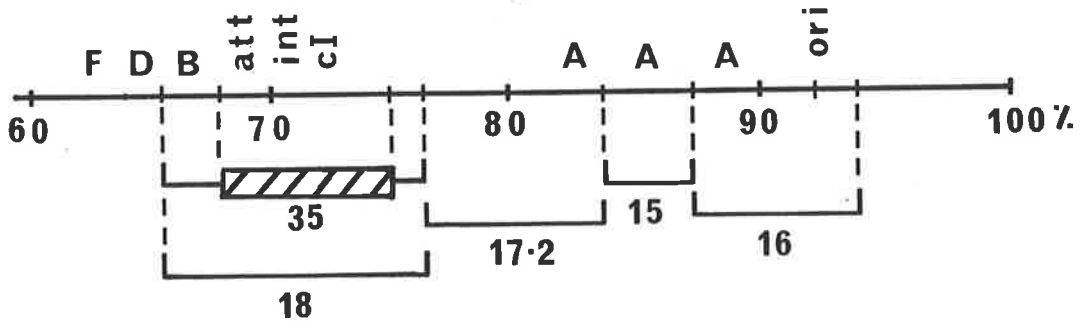
Transcription which was dependent on phage protein synthesis was investigated by adding chloramphenicol immediately prior to induction, to halt protein synthesis. The interval in pEC35 (Fig 1.3) was the only region that showed high transcriptional activity in the presence of chloramphenicol. This interval spans the region 65.4% to 76.3% (λpEC18, Fig. 1.3) but contains a deletion (Δ1) from 67.9% to 74.9%. Having concluded previously that B gene transcription (the B gene is encoded in the interval 65.4% to 67.9%) was dependent on functional A protein, Finnegan and Egan (1981) also concluded that transcription in the presence of chloramphenicol was mainly from the interval 74.9% to 76.3%. From these results Finnegan and Egan (1981) postulated that 186 encoded a gene X (at

Fig. 1.3. Clone map of the early region of phage 186

The location and extent of PstI fragments of 186 DNA which were cloned into the plasmid pBR322 and used for in vivo mRNA hybridisation studies (Finnegan and Egan, 1981) is shown.

Plasmid pEC35 carries a deletion (Δ1) of the int-cI region (indicated by the shaded box) from 67.9% to 74.9% (later shown to be from 67.9% to 74.0%, see Section 4.2.7). Plasmid clone pEC18 which contains the PstI fragment from 65.5% to 76.3% (later shown to be from 65.5% to 77.4%, see Section 3.2) was found to be unstable (Finnegan and Egan, 1981). The entire plasmid was subsequently cloned into λ vector λ762 and this clone (λpEC18) was used for in vivo mRNA hybridisation studies (Finnegan and Egan, 1981).

All known alleles of the replication gene A map between 83.8% and 92.0% of the genome (Finnegan and Egan, 1979).


F D B att int cl A A A ori
60
70
80
90
100 %
35
17·2
15
16
18

about 76%), which was required for further rightward transcription into the intervals pEC17.2 and pEC15 (Fig. 1.3). Gene X could function either by initiating new transcription or by antiterminating its own transcript in a manner analogous to the action of the λ N gene-product. Transcription of the interval 74.9% to 76.3% was under cI repressor control and allowed the operator (O) to be placed between the cI gene and X.

These studies were constrained by a lack of knowledge of the exact gene content of the cloned fragments and the exact extent of the deletion in the interval pEC35. Although these experiments provided a crude pattern of phage 186 *in vivo* transcripts, further studies were required to provide a more accurate description of 186 *in vivo* transcripts.

Previous discussion of the transcriptional controls of phage λ (Section 1.2) indicated that a knowledge of the precise location, size, gene content and direction of transcription of individual RNA transcripts is an essential prerequisite to understanding more precisely how transcription is controlled. In phage 186, the early control region has the potential for encoding more genes than have been detected by genetic studies (Section 1.3.1) and these genes must be identified to facilitate studies on the control of early gene expression.

1.6 DNA ANALYSIS

1.6.1. DNA Sequencing

Undoubtedly the two developments that have led to a significant acceleration in the analysis of gene expression have been recombinant DNA technology and rapid DNA sequencing techniques. Two methods are currently widely used to determine DNA sequences; the chemical cleavage technique of Maxam and Gilbert (1977) and the dideoxynucleotide

chain termination technique of Sanger et al., (1977). The power of the latter technique was greatly enhanced with the introduction of the single-stranded phage M13 as a cloning vector (Messing et al., 1977; Sanger et al., 1980).

1.6.2 Computer Analysis

Sufficient DNA sequence data of known prokaryotic genes has accumulated to provide a statistical basis for several methods of predicting genes (protein-coding regions) from the DNA sequence alone (Fickett, 1982; Shepherd, 1981; Staden and McLachlan, 1982, Staden, 1984a). The codon usage method of Staden and McLachlan (1982) was used successfully to locate the genes of phage λ (Sanger et al., 1982) by comparing the distribution of codons of known λ genes (e.g. gene J as a standard) with those observed by scanning the sequence. Assuming the codon usage of all genes in the sequence was similar, reading frames which show a codon distribution most like that of the standard were considered to be protein-coding regions.

Signals important in gene expression such as transcription promoters, translational initiation sites (Rosenberg and Court 1979; Siebenlist et al., 1980)) and rho-independent transcription terminators (Adhya and Gottesman, 1978) can be located by computer programs that search for DNA sequences which have common features with those sequences known to act as signals for gene expression (Staden, 1984b; Staden, 1984c; Brendel and Trifinov, 1984). A powerful method of identifying genes is to use a combination of searching the DNA sequence for gene content (e.g. codon usage) and for signals important in gene expression. This approach was incorporated into the ANALYSEQ program (Staden, 1984b) and used successfully to identify λ genes (Sanger et al., 1982).

1.7 RNA ANALYSIS

1.7.1. Mapping in vivo RNA Transcripts by RNA/DNA Hybridisation

A powerful method for mapping *in vivo* RNA transcripts is to hybridise uniformly labelled RNA to defined DNA segments using the nitrocellulose filter hybridisation techniques of Nygaard and Hall (1963) and Gillespie and Spiegelman (1965). These basic procedures are still widely used to study *in vivo* transcription patterns in prokaryotes (reviewed in Hall *et al.*, 1980).

An accurate *in vivo* transcription pattern for phage λ was determined in this manner, using λ DNA segments defined by deletion mutants, hybrid phages and separated strands (reviewed in Bφvre *et al.*, 1971) or sheared DNA halves (Lindqvist and Bφvre, 1972; Hershey *et al.*, 1965). Multistep RNA-DNA hybridisation experiments, where radioactive RNA was successively hybridised to and eluted from various DNA fragments to enrich for a particular mRNA species, increased the discriminatory power of these types of techniques, (reviewed in Bφvre and Szybalski, 1971). The potential of RNA-DNA hybridisation experiments has been greatly enhanced by the use of recombinant DNA technology which allows the construction of clones containing defined DNA restriction fragments. Plasmid clones of 186 DNA were used to provide the general *in vivo* RNA transcription pattern of this phage (Finnegan and Egan, 1981).

The pattern of *in vivo* RNA transcription indicates which regions of the genome are transcriptionally active during the course of phage infection and is an important first step in understanding the control of transcription. A knowledge of the size of individual RNA transcripts is also required.

1.7.2 Physical Characterisation of RNA Transcripts

Physical characteristics of RNA transcripts can be determined by sedimentation of the RNA species through sucrose density gradients or by determining the electrophoretic mobility of the RNA species. (Kourilsky et al., 1968; Summers, 1969; Studier, 1973). Major problems with these techniques are the presence of E. coli RNA species which obscure phage-encoded RNA transcripts, and rapid breakdown of mRNA molecules. The former problem can be resolved by enriching for phage mRNA species by multistep RNA-DNA hybridisation to enrich for phage specific mRNA (Section 1.7.1) or by inhibiting E. coli RNA synthesis prior to phage infection by high doses of UV-irradiation (Studier and Maizel, 1969).

A different approach to detecting and sizing individual RNA transcripts is to fractionate the RNA on an agarose gel and transfer the RNA to a nitrocellulose membrane (Thomas, 1980). RNA species are then detected by hybridisation with radioactive DNA probes (Northern analysis). This technique detects mRNA transcripts accumulated during the course of phage infection so that the products of RNA processing and breakdown are also be detected. Despite this disadvantage, the technique was used to identify phage 186-specific RNA transcripts. A library of single-stranded M13 clones could be used to construct very specific hybridisation probes to identify the 186 RNA transcripts.

1.7.3 RNA Sequence Analysis

The most definitive approach to RNA analysis is to determine the nucleotide sequence of the RNA species of interest, and map it to the DNA sequence.

Direct RNA sequencing techniques rely on the isolation of

the intact RNA species. RNA can be uniformly labelled *in vivo* with ^{32}P, prior to its isolation, and the sequence can be determined by two-dimensional fingerprint analysis (Barrell, 1971; Schmeissner *et al.*, 1980; Court *et al.*, 1980). Alternatively the RNA species can be labelled at either terminus, after its isolation, and sequenced by the direct enzymatic sequencing technique of Donis-Keller *et al.* (1977).

Recombinant DNA technology allowed the development of techniques for rapid RNA sequencing that did not necessarily require isolation of the mRNA species of interest (reviewed by Godson, 1980). These techniques rely on a knowledge of the DNA sequence and the approximate position of the RNA transcript on the genome, to allow the construction of radioactive DNA primers which hybridise specifically to the RNA species of interest. The sequence from the priming site can then be determined by the dideoxynucleotide chain termination technique, using AMV reverse transcriptase (reviewed by Godson, 1980). By judicial selection of a DNA primer, the 5'-terminus of the RNA transcript can be identified.

A simpler approach to identify the 5'-terminus of the transcript is to extend the radioactive DNA primer in the presence of all four dNTP's and accurately determine the size of the extension product by its migration relative to a DNA sequencing ladder (McKnight *et al.*, 1981). An alternative method is the S1 nuclease mapping technique (Berk and Sharp, 1977) which has been used successfully to map the 5'-termini of several phage P2 late mRNA species, to the P2 DNA sequence (Christie and Calendar, 1983, 1985). Primed synthesis methods were used to locate the 5'-termini of 186 *in vivo* transcripts on the DNA sequence.

1.8 RATIONALE AND AIM

Studies aimed at characterising *in vivo* (this work) and *in vitro* transcripts (Pritchard, 1984; Pritchard and Egan, 1985) were initiated in conjunction with the determination of the 186 DNA sequence, to allow definitive mapping of the transcripts and identification of the potential genes.

Characterisation of the *in vitro* transcripts revealed transcripts initiated by unmodified host RNA polymerase, in the absence of phage protein synthesis, and were representative of the initial phage transcripts made upon infection. The *in vivo* transcripts would be those produced during the normal course of infection. This type of experimental approach has been crucial to the elucidation of λ gene control mechanisms.

A 3.3 kb *Pst*I(65.5%-76.3%) fragment was chosen for DNA sequencing, since it conveniently spanned the early control region (Finnegan and Egan, 1979) and fractionated well away from other *Pst*I fragments (Saint and Egan, 1979). This fragment is in fact rich in control functions including the genes required for integration (*int*), late gene control (*B*), and maintenance of repression (*cI*). The fragment also includes two important sites, the prophage attachment site (*att*), which plays a pivotal role in site-specific recombination, and the operator region which may reveal important insights into the control of the lysis-lysogeny decision in phage 186. Finally, the *Pst*I(65.5%-76.3%) fragment is thought to contain the primary control gene *X*, which is predicted to map around 76% on the 186 chromosome and is required to foster transcription to the right of 76%.

The aim of this work was to determine the DNA sequence of the 3.3 kb *Pst*I(65.5%-76.3%) DNA fragment, identify the potential

genes on the fragment, and where possible, to characterise the *in vivo* RNA transcripts from this region. A broader aim of this work was to investigate the potential of phage 186 for displaying new mechanisms by which genes can be controlled.

SECTION 2

MATERIALS AND METHODS

2. MATERIALS AND METHODS

2.1 BACTERIAL STRAINS

Bacterial strains used during the course of this work are described in Table 2.1.

2.2 BACTERIOPHAGE STRAINS

2.2.1 186 Strains

186cItsp: a heat inducible mutant with a temperature sensitive cI repressor (Woods and Egan, 1974; Baldwin et al., 1966).

186cIam53vir1: a virulent phage able to grow on a 186 lysogen, isolated as a spontaneous mutant in a stock of the phage 186cIam53 (Woods, 1972).

186cItspDam14vir2: has an amber mutation in the tail gene D, allele 14. The vir2 mutation appeared as a spontaneous mutant in a stock of 186cItspPam16 (Woods, 1972) and was recombined into 186cItspDam14 (Hocking and Egan, 1982a).

186cItspDam23vir2: has an amber mutation in the tail gene D, allele 23. The vir2 mutation was recombined into this strain as described for 186cItspDam14vir2

186cItspBam17vir3: has an amber mutation in the late control gene B, allele 17 (Hocking, 1977). The vir3 mutation was isolated as a spontaneous mutant in a stock of the phage 186cItspBam17 (J.B. Egan, unpublished).

2.2.2 P2 Strains

P2vir22ogr52: has a 5% deletion from 72.2% to 77.2% and a 0.5% insertion at 72.2% (Chattoraj

TABLE 2.1 - BACTERIAL STRAINS

Collection No.	Strain or other No.	Genotype	Relevant Character	Origin or reference
E251	W3350	F^- galK galT strA	Su^-	This laboratory by transduction of the strA allele from CGSC 4214 into W3350
E252	E251(186)	E251(186cItsp)	Su^- lysogen of 186cItsp	Finnegan, (1979)
E264	E251(186Bam17)	E251(186cItspBam17)	Su^- lysogen 186cItspBam17	Finnegan, (1979)
E508	C600	tonA supE44 thr leu thi	Su^+	Appleyard, (1954)
E574	C600	C600(186cItsp)	Su^+ lysogen of 186cItsp used for fermentation preparations of phage (2.14.3)	This laboratory from 186cItsp described by Baldwin et al. (1966).
E536	W3350	F^- galK galT strA	Su^-	Campbell, (1965)
E402	CGSC 5112 (Hy5)	F^+ thi thyA thyR uvrA (Hy5)	Hy5 lysogen for phage preparations	Rupp et al. (1971)
E2234		E536 carrying plasmid pEC35	Source of pEC35 DNA	Finnegan, J.
E0605	JM101	lac pro supE44 thi F' traD36 proAB $lacI^q$ ZΔM15	Host for M13 infection	Messing, (1979)

and Bertani, 1980): ogr52 is a mutation in the ogr gene permitting growth on an E. coli gro109 mutant (Sunshine and Sauer, 1975).

2.2.3 M13 Strains

M13mp7, M13mp8, M13mp9 (Messing et al., 1981; Messing and Vieira, 1982) were used to prepare vector DNA for cloning DNA restriction fragments to be sequenced. An M13mp9 clone containing the 3.0 kb PstI-Bg1II(62.2%-77.2%) fragment of P2vir22ogr52 DNA was a kind gift from Dr. M. Pritchard.

2.2.4 P2.186 Hybrid Strains

Hy5 is a P2.186 hybrid phage isolated by Bradley et al. (1975) containing 0% to 70.6% of the P2 genome and 65.4% to 100% of the 186 genome (Younghusband et al., 1975) with the crossover point in the 186 E gene and the P2 D gene (Hocking and Egan, 1982d).

2.3 PLASMID CLONES

pEC35: 186 PstI restriction fragment (65.5%-76.3%) containing a deletion (Δ1) from 67.9% to 74.9%, cloned into the PstI site of pBR322 (Finnegan and Egan, 1981). After DNA sequencing analysis (Section 4.2.7), the percentage positions were revised and pEC35 will be referred to as a plasmid clone of the PstI(65.5%-77.4%)

fragment containing a deletion from 67.9% to 74.0%.

MEC18: 186 PstI(65.5%-77.4%) fragment from Hy5 DNA cloned into the PstI site of M13mp7 vector DNA. Two isolates MEC18(L) and MEC18(R) containing the insert in either orientation were obtained, and the insert size was confirmed by restriction analysis (2.19.3). Single-strand DNA of each isolate was prepared and sequenced to establish the orientation of the insert (2.20). Single-stranded clones MEC18(L) and MEC18(R) were used to test the orientation of M13 clones (Section 2.19.8(c)).

2.4 ENZYMES

Restriction endonucleases: New England Biolabs or Bethesda Research Labs.

E. coli DNA polymerase I (Klenow fragment) : Boehringer Mannheim and Biotechnology Research Enterprises of South Australia (BRESA).

T4 DNA ligase: Boehringer Mannheim.

Calf intestinal phosphatase: Sigma Chemical Co. Purified according to the procedure of Efstratiadis et al. (1977), was a gift from Dr. R.H. Symons.

Avian myeloblastosis virus (AMV) reverse transcriptase : Life Science Inc., Florida.

E. coli RNA polymerase (holoenzyme): Boehringer Mannheim

Lysozyme: Sigma Chemical Co.

Proteinase K: Boehringer Mannheim

RNAse A: Sigma Chemical Co. Stock solutions were heated at 80°C for 10 to 15 minutes to inactive DNAses.

E. coli DNAse I: Boehringer Mannheim

2.5 RADIOCHEMICALS

Radiochemicals d[α-^{32}P]CTP and d[α-^{32}P]ATP of specific activity 1700 Ci/mmol and [α-^{32}P]GTP of specific activity 1000 Ci/mmol were initially prepared by Dr. R.H. Symons and subsequently purchased from BRESA.

2.6 CHEMICALS

All chemicals were of analytical grade or of the highest purity available unless otherwise specified.

Polyethylene glycol (PEG) 8000: for phage preparations and general use was from Sigma Chemical Co. M13 phage preparations for sequencing (2.19.6) were prepared using PEG 8000 from B.D.H. Labs.

Urea: Sigma Chemical Co.

Cesium Chloride (CsCl): Bethesda Research Labs.

Phenol: AR grade, was redistilled and stored under nitrogen in the dark at -15°C, B.D.H. Labs.

Bacto-tryptone, yeast extract and Bacto-agar: Difco Labs., U.S.A.

Tetracycline: Upjohn Pty. Ltd., a kind gift.

Amine A: Humpko Sheffield, U.S.A.

Sodium dodecyl sulphate (SDS): Sigma Chemical Co.

Sodium azide (NaN_3): Ajax Chemicals Ltd.

Trizma base and Tris 7-9: Sigma Chemical Co.

Ethylenediaminetetraacetic acid (EDTA): Sigma Chemical Co., disodium salt.

Piperazine-N, N'-bis-(2-ethane sulphonic acid) (PIPES): Sigma Chemical Co.

Acrylamide: Sigma Chemical Co.

N, N'-methylene-bis-acrylamide (bis): Sigma Chemical Co.

Ammonium persulphate (APS): May and Baker. Stock solutions at 25% (w/v) in water, were kept at $4^{o}C$.

Agarose: Sigma Chemical Co.

Low gelling temperature agarose (LGT): Bethesda Research Labs.

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

Isopropyl-β-D-thiogalactopyranoside (IPTG): Sigma Chemical Co. Stock solutions were used at 24 mg/ml in water, and kept at $-15^{o}C$.

5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (BCIG): Sigma Chemical Co. Stock solutions at 20 mg/ml in dimethyl formamide, were kept at $-15^{o}C$.

Ribonucleoside triphosphates (NTP): Sigma Chemical Co. Stock solutions at 20 mM (prepared in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) were kept at $-15^{o}C$.

Deoxyribonucleoside triphosphates (dNTP): Sigma Chemical Co. Stock solutions at 20 mM (prepared in 5 mM Tris-HCl, pH 8.0, 0.1mM EDTA) were kept at $-15^{o}C$.

Dideoxyribonucleoside triphosphates (ddNTP): Sigma Chemical Co.

	Stock solutions at 20 mM (prepared in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) were kept at -15°C.
Glyoxal:	Technical grade obtained from B.D.H. Labs. Aust. De-ionised (2.18.4(e)) and stored in the dark at -80°C for long term storage, otherwise stored at -15°C.
Formamide:	B.D.H. Labs. Aust. De-ionised (2.18.4(e)) and stored in the dark at -15°C.
Mixed bed resin (508-X8(D)):	Bio-Rad Labs.
Sequencing Primer:	(17-mer ; 5'-GTAAAACGACGGCCAGT-3') was purchased from New England Biolabs and BRESA.
Dithiothreitol (DTT):	Sigma Chemical Co. Stored as a 1 M solution in water, in the dark at -15°C.
Ethidium bromide (EtBr):	Sigma Chemical Co. Stored as a 10 mg/ml solution in water, at 4°C.
Acridine orange:	Sigma Chemical Co.
Ethanol:	Redistilled before use and stored at -15°C. RNAse-free ethanol was obtained by sterilising the distillation apparatus and collection bottles in a 110°C oven overnight before use.
Calf thymus DNA:	Sigma Chemical Co. Sonicated, heat denatured and stored as a 10 mg/ml solution in 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA at -15°C.
E. coli tRNA:	Sigma Chemical Co. Extracted three times with phenol/water before use and kept as a 10 mg/ml solution in water, at -15°C.

Bovine serum albumin (BSA): Sigma Chemical Co. Acetylated before use to remove nucleases according to the procedure of Gonzalez et al. (1977) and kept as a 2 mg/ml solution in water at -15°C. Gift from Dr. M. Pritchard.

2.7 MEDIA

2.7.1 Liquid Media

All media and buffers were prepared in glass-distilled water and were sterilised by autoclaving for 25 minutes at 120°C and 120 kPa.

L broth: 1% Bacto-typtone, 0.5% yeast extract, 1% NaCl, pH 7.0.

2x YT broth: 1.6% Bacto-typtone, 1% yeast extract, 0.5% NaCl, pH 7.0.

M13 minimal medium: 1.05% K_2HPO_4, 0.45% KH_2PO_4, 0.1% $(NH_4)_2SO_4$, 0.05% Na_3 citrate.$2H_2O$ and water to 1000 ml. This solution was autoclaved, cooled to 45°C and the following added from separately prepared sterile solutions; 10 ml of 20% glucose, 0.8 ml of 1M $MgSO_4$, 0.5 ml of 1% thiamine-HCl.

2.7.2 Solid Media

Z plates: 1% Amine A, 1.2% Bacto-agar, 0.5% NaCl, pH 7.2.

M13 minimal plates: 1.5% Bacto-agar was added to M13 minimal media.

Antibiotic plates: Z plates were supplemented with the antibiotic tetracycline at 20 mg/ml for the growth of the pBR322 plasmid

clones used in this work. Plates were poured from 30 ml of the appropriate medium, dried overnight and stored at $4^{o}C$.

Soft agar overlay: 0.7% Bacto-agar, pH 7.0.

YT soft agar overlay: 0.8% Bacto-tryptone, 0.5% yeast extract, 0.5% NaCl, 0.7% Bacto-agar, pH 7.0.

2.8 BUFFERS

20x SSC: 0.15 M NaCl, 0.015 M Na_3 citrate, pH 7.4. 0.89 M Tris, 0.89 M boric acid, 2.7 mM EDTA, pH 8.3.

10x TAE: 0.4 M Tris-acetate, 0.2 M Na acetate, 10 mM EDTA, pH 8.2.

10x TE: 0.1 M Tris-HCl, pH 8.0, 1 mM EDTA.

10x TM: 0.1 M Tris-HCl, pH 8.0, 0.1 M $MgCl_2$. TM used for preparation and storage of phage (2.13, 2.14, 2.15) was 10 mM Tris-HCl, pH 7.1, 10 mM $MgSO_4$.

2.9 MOLECULAR WEIGHT MARKERS

2.9.1. DNA Molecular Weight Markers

HinfI digest of pBR325 DNA at 400 ng/ul. HindIII digest of phage lambda DNA at 400 ng/ul, HaeIII digest of ϕX174 DNA at 700 ng/ul, were from Bethesda Research Laboratories. Sizes are marked on the appropriate figures.

2.9.2. RNA Molecular Weight Markers

Cucumber mosaic virus (CMV) RNA was a gift from Dr. R.H. Symons. RNA 1 (3387 b), RNA 2 (3035 b), RNA 3 (2193 b), RNA 4 (1027 b). Sizes were given in Rezain and Williams, (1984). Ribosomal RNA markers were E. coli 23S RNA (2904 b) and E. coli 16S RNA (1541 b) (Brosius et al., 1978; Brosius et al., 1980).

2.10 MISCELLANEOUS MATERIALS

Fuji Rx medical X-ray film: Fuji Photo Film Co.

Positive/Negative Land Pack film: Polaroid

Nitrocellulose: Schleicher and Schuell BA85 (0.45 um)

Ultrafiltration membrane filters: Millipore (0.45 um)

Dialysis membrane (18/32): Union Carbide

2.11 STORAGE OF BACTERIA AND BACTERIOPHAGE

Bacterial stocks for short term storage were maintained on Z plates at $4^{o}C$ except JM101 which was maintained on M13 minimal plates. Long term storage of bacterial cultures was by freezing at $-80^{o}C$, after addition of an equal volume of 80% glycerol.

Low titre stocks of M13 recombinant phage were maintained in 2x YT broth at $-15^{o}C$. High titre 186 phage stocks prepared by CsCl block gradient centrifugation (2.15), were dialysed three times against one litre of TM and stored at $4^{o}C$. Low titre 186 phage stocks were passed through Millipore membrane filters and stored at $4^{o}C$.

2.12 GROWTH OF BACTERIAL CULTURES

Stationary phase bacterial cultures were prepared by inoculating broth with a single colony of bacteria from a plate stock, or for JM101, a loopful of bacteria directly from the glycerol stock, and incubating overnight in capped flasks at $30^{o}C$ or $37^{o}C$ in a New Brunswick gyrotary water bath.

Log phase cultures and indicator bacteria were prepared by diluting a fresh stationary culture fifty-fold into sterile broth and incubating with aeration in a gyrotary water bath at $30^{o}C$ or $37^{o}C$, until the required cell density was reached (usually A_{600} = 0.8, 5 x 10^{8} cells/ml). Cell density was measured by observing

the A_{600} using a Gilford 300 T-1 spectrophotometer. Indicator bacteria were chilled and kept on ice until required.

2.13 TITRATION OF PHAGE STOCKS

Phage lysates were assayed by plating 0.1 ml of the appropriate dilution in TM buffer, 0.2 ml of log phase indicator bacteria (2.12) and 3 ml melted soft agar overlay on Z plates. The agar was allowed to solidify and the plates were inverted and incubated overnight at 37°C. Plaques were scored on the following day.

2.14 PREPARATION OF PHAGE STOCKS

2.14.1 Low Titre Phage Stocks

Low titre stocks of 186 phage were prepared by heat induction or liquid infection as described by Hocking and Egan (1982a).

2.14.2 High Titre Phage Stocks by Heat Induction of 500 ml Cultures

A fresh stationary phase culture was grown at 30°C, diluted fifty fold into 500 ml L broth and incubated at 30°C with aeration to an A_{600} = 0.8. The culture was transferred to a 45°C water bath, shaken by hand for 15 minutes, and then returned to a 39°C water bath and shaken for a further three to four hours.

Bacterial debris was removed by centrifugation (8,900 g, 4°C, 20 minutes) and the supernatant was decanted. NaCl and PEG 8000 were added to a final concentration of 0.5 M and 10% (w/v) respectively and precipitation allowed to proceed overnight at 4°C. The precipitate was collected by centrifugation (8,900 g, 4°C, 20 minutes), resuspended in 8 ml of TM and then purified by CsCl block gradient centrifugation (2.15).

2.14.3 High Titre Phage Stocks by Heat Induction of Eight Litre Cultures

A fresh stationary phase culture was diluted sixty-fold into eight litres of L broth and incubated at 30°C in a New Brunswick Magnaferm Fermentor MA 100 at 400 rpm agitation speed and with

8,000 cc/minute of air circulating through the culture. Heat induction was achieved by raising the temperature to 39°C and the culture incubated for a further three to four hours. Harvesting the culture and precipitation of the phage particules was as described in 2.14.2.

2.14.4 High Titre Phage Stocks by Liquid Infection

A fresh overnight culture (2.12) of E508 was diluted one hundred fold into 500 ml L broth (prewarmed to 37°C) and then incubated with aeration to an A_{600} = 0.6. The culture was infected (moi = 0.1) with phage from a low titre stock (2.14.1), incubation continued at 37°C and 1 M $MgSO_4$ added to a final concentration of 0.01 M at the onset of lysis. Addition of $MgSO_4$ was found to increase the phage yield, however the reason for this increase was not known. When lysis was complete, 1 ug/ml DNAse I was added and then solid NaCl dissolved into the lysate to give a concentration of 25 ug/ml. The lysate was chilled for 20 minutes on ice, bacterial debris removed by centrifugation (8,900 g, 20 minutes, 4°C) and the supernatant decanted. PEG 8000 (10% w/v) was dissolved into the supernatant and the precipitation was allowed to proceed for 90 minutes on ice. The precipitate was collected by centrifugation (8,000 g, 20 minutes, 4°C) and the phage pellets were allowed to resuspend overnight in 8 ml TM. Phage were purified by CsCl gradient centrifugation as described in 2.15.

2.15 CsCl DENSITY GRADIENT CENTRIFUGATION

2.15.1 CsCl Block Density Gradient

Two CsCl solutions of density 1.6 g/ml and 1.35 g/ml were prepared in sterile TM and used to form a block gradient by adding 3 ml of the 1.35 g/ml solution and underlaying it with 1.5 ml of the 1.6 g/ml solution, in a 10 ml polycarbonate Oakridge tube. The high titre phage suspension (2.14.2) in TM was carefully layered on top of the gradient and the tube centrifuged for 90 minutes

at 45,000 rpm, 8°C in a Beckman Ti 50 rotor.

The opaque band was collected by piercing the bottom of the tube and an equal volume of saturated CsCl solution was added to the phage suspension. A block gradient of 1.5 ml 1.6 g/ml density and 2.5 ml 1.35 g/ml density was underlayed with the phage suspension, topped with paraffin oil, and centrifuged as for the first block gradient. Again the opaque phage band was collected, dialysed three times against one litre of TM, and stored at 4°C.

2.15.2 CsCl Equilibrium Density Gradient

An alternative to CsCl block gradient centrifugation (2.15.1) for the preparation of phage particles was CsCl equilibrium density gradient centrifugation. The high titre phage suspension (2.14.2) was diluted with TM and the required amount of CsCl dissolved into the solution to give a density of 1.45 g/ml. After centrifugation to equilibrium (32,000 rpm, 24 hours, 8°C, Beckman Ti 50 rotor), the opaque phage band was collected by piercing the tube from the side, dialysed three times against one litre of TM, and stored at 4°C.

Plasmid DNA was purified by dissolving the precipitated nucleic acids in 8 ml TE (2.17.2), to which 8 gm of CsCl and 200 ul of 10 mg/ml solution of EtBr was added. After centrifugation to equilibrium (45,000 rpm, 42 hours, 15°C, Beckman Ti 50 rotor), the bands were visualised under subdued fluorescent light and of the two bands, the lower band containing the replicative form of plasmid DNA was collected by piercing the tube from the bottom. EtBr was removed by three extractions with isopropanol equilibrated with 5 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA. The DNA solution was dialysed three times against one litre of TE and stored at 4°C.

2.16 PHENOL EXTRACTION OF DNA

2.16.1 Phenol Equilibration and Storage

Buffer equilibration of phenol was carried out by mixing 50 ml redistilled phenol with 50 ml of 1 M Tris-HCl, pH 8.0, and 5 mg of 8-hydroxy quinoline. The phases were allowed to separate, the phenol phase was taken and mixed with 50 ml of TE, and allowed to stand until the phases separated. The aqueous phase was removed and the phenol phase was again equilibrated with TE. Phenol equilibrated with TE in this manner was stored under TE and kept frozen in 50 ml aliquots at -20^{o}C until required.

2.16.2 Extraction of Bacteriophage DNA

A high titre phage stock (10^{13} pfu/ml, 2.15) was diluted to 0.9 ml in TE and 0.1 ml of 100 mM Tris-HCl, pH 8.0, 100 mM NaCl, 100 mM EDTA, 5.0% SDS was added, followed by 0.1 mg of Proteinase K. After incubation at 37^{o}C for 60 minutes, the solution was again diluted to 5 ml with TE and an equal volumme of TE-equilibrated phenol (2.16.1) added. The mixture was gently shaken for 5 minutes and the phases separated by centrifugation in 50 ml polypropylene Oakridge tubes at 3,000 g, 4^{o}C for 5 minutes. The aqueous phase was withdrawn and re-extracted at least twice with an equal volume of TE-equilibrated phenol. NaCl was added to the pooled aqueous phases to a final concentration of 0.2 M, followed by 2.5 volumes of ethanol and DNA was allowed to precipitate overnight at -20^{o}C. DNA was collected by centrifugation (10,000 g, 4^{o}C, 20 minutes), the pellet washed in 70% ethanol/TE (v/v) and rinsed in redistilled ethanol. The pellet was dried *in vacuo* for 10 minutes, dissolved in TE and stored at 4^{o}C.

The concentration and purity of the DNA was determined by obtaining the spectrum of absorbance over the range 230-340 nm

with a Varian Superscan 3 ultra-violet/visible spectrophotometer. $A_{260/280}$ and $A_{260/230}$ ratios were greater than 1.8 for all DNA used.

2.16.3 Phenol Extraction of DNA solutions

DNA solutions were mixed with a half volume of TE-equilibrated phenol in Eppendorf tubes, mixed for 5 minutes and centrifuged (10,000 g, 3 minutes, room temperature). The aqueous phase was withdrawn and the process repeated until no proteinaceous material was visible at the interface of the aqueous and phenol phases. DNA was ethanol precipitated from the aqueous phase and dissolved in TE as described for the extraction of bacteriophage DNA (2.16.2), except addition of Na acetate, pH 5.2, to 0.3 M was used instead of NaCl.

2.17 PLASMID DNA PREPARATION

2.17.1 Analytical Preparation

The following procedure, based on the method of Birmboin and Doly (1979), gave sufficient plasmid DNA for restriction analysis from only 1 ml of a stationary phase, plasmid containing culture.

Bacteria were pelleted by centrifugation (10,000 g, 2 minutes, 4°C) and the pellet resuspended in 100 ul of lysis buffer (25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 15% sucrose to which 1 ul of a 5 mg/ml freshly prepared lysozyme solution was added). After 40 minutes on ice, 200 ul of 0.2 N NaOH, 1% SDS was added and the tube placed on ice for a further 10 minutes. After addition of 150 ul of 3 M Na acetate, pH 5.2, the tube was placed on ice again for 40 minutes. Chromosomal DNA was pelleted by centrifugation (10,000 g, 10 minutes 4°C) and plasmid DNA in the supernatant was precipitated by addition of two volumes of ethanol. The pellet was washed in 70% ethanol (v/v in TE), dried *in vacuo* for 10 minutes and redissolved in 39 ul water. RNAse A (1 ul of a 5 mg/ml solution, 2.4) was added and incubation was for 20 minutes at 37°C. Restriction

analysis (2.19.3) was carried out without any further treatment to the plasmid DNA solution.

2.17.2 Preparative

A stationary culture of plasmid containing cells was prepared (2.12) in 10 ml of L broth containing antibiotic (20 ug/ml tetracycline for pEC35) and used to inoculate 800 ml of L broth. The culture was grown with aeration to A_{600} = 1.0, and then 160 mg of chloramphenicol added and amplification of the plasmid was allowed to proceed at 37°C overnight. Cells were pelleted by centrifugation (5,000 g, 10 minutes, 4°C), resuspended in 7.5 ml of 25% sucrose, 50 mM TrisHCl, pH 8.0, and then 2.0 ml of a freshly prepared 10 mg/ml solution of lysozyme was added. The tube was placed on ice for 5 minutes and 3 ml of 0.25 M EDTA, pH 8.0, added. After gentle mixing the tube was again placed on ice for 5 minutes. Detergent solution (12 ml of 1% Brij 58 and 0.4% Na deoxycholate (w/v) in 50 mM Tris-HCl, pH 8.0, 25 mM EDTA) was added, mixed gently, and the solution placed on ice for a further 10 minutes. After centrifugation (40,000 g, 30 minutes, 4°C), the supernatant was removed and plasmid DNA in the supernatant was collected by ethanol precipitation. The pellet was dissolved in 8 ml TE and plasmid DNA purified by CsCl equilibrium gradient centrifugation (2.15.2).

2.18 GEL ELECTROPHORESIS

2.18.1 Agarose Gel Electrophoresis of DNA

Agarose gel solution (1% w/v agarose in 1 x TAE) was stored at 65°C. Analytical work was carried out on minigels prepared by pouring 9 ml of gel solution onto a 7.5 cm x 5.0 cm glass microscope slide, with the appropriate comb set in place. Glycerol loading buffer (5% glycerol, 10 mM EDTA, 0.04% bromophenol blue) was added to the samples and electrophoresis was at 200 mA.

Preparative gels were poured in a glass sandwich (20 x 20 x 0.15 cm) and run in a vertical position or poured into a perspex tray (14 x 11 x 0.3 cm) and run horizontally, submerged in 1x TAE.

DNA was visualised by staining with EtBr (0.0004% w/v in 1x TAE) and photographed under short wavelength UV-light. Approximate concentrations of DNA solutions were determined by comparing the intensity of the EtBr stained bands with the intensity of HindIII digested lambda DNA (2.9.1) bands of known concentration.

2.18.2 Agarose Gel Electrophoresis of RNA

Agarose gels for RNA fractionation were 1.8% agarose (w/v) in 10 mM Na phosphate, pH 6.5. The gel solution was poured into a perspex tray (14 x 11 x 0.3 cm) and gels were run horizontally, submerged in 10 mM Na phosphate, pH 6.5. The buffer was recirculated every 15 minutes by hand, or continuously by a peristaltic pump. Glycerol loading buffer (5% glycerol, 10 mM Na phosphate, pH 6.5, 0.04% bromophenol blue) was added to the samples and electrophoresis was at 30 mA.

Gels were stained with EtBr (0.0004% w/v in 10 mM Na phosphate, pH 6.5) and photographed under short wavelength UV-light.

2.18.3 Extraction of DNA from Agarose

DNA to be extracted from agarose for the purpose of cloning, was detected by staining with acridine orange (0.001% w/v in 1x TAE) for 10 minutes, destained in 1x TAE for at least 15 minutes and the bands were visualised under fluorescent light. Two methods were used to extract the DNA from the agarose slice.

2.18.3(a) Extraction of DNA from Low Gelling Temperature (LGT) Agarose

LGT agarose gel solution (1.5% w/v in 1x TAE) was cooled to 37^{o}C before pouring (2.18.1) and electrophoresis was at

100 mA. After staining with acridine orange the DNA bands, visualised by eye, were excised with a flamed scalpel blade and two volumes of 50 mM Tris-HCl, pH 8.0, 0.5 mM EDTA added. The gel slice was then melted at 60°C for 30 to 45 minutes, transferred to a 37°C water bath and extracted with phenol as described in 2.16.3, except the tubes were maintained at 37°C during the extraction.

2.18.3(b) Electro-elution from Agarose Slices

Dialysis tubing (2.18.4(f)) was slit to form a sheet and pushed into a 0.5 x 4.0 cm slot cut into the centre of a 8.0 x 8.0 x 0.15 cm polypropylene sheet, to form a trough. The slice containing the desired DNA fragment was placed into the trough with enough 1x TAE to just cover the slice (usually 400 ul). The trough was placed perpendicularly to the path of electric current in a horizontal gel apparatus, and electrophoresis was at 100 mA for 10 to 20 minutes. The current was reversed for one minute and the buffer surrounding the slice was removed, after using a portion of it to rinse the trough walls and remove any DNA sticking to the walls. DNA was precipitated with ethanol after addition of 5 ug of E. coli carrier tRNA and one tenth volume of 3M Na acetate pH 5.2 (2.16.3).

2.18.4 POLYACRYLAMIDE GEL ELECTROPHORESIS

2.18.4(a) Non-denaturing Gels

A 50% gel stock (acrylamide:bis 30:0.8) was prepared by dissolving 243.5 gm acrylamide and 6.5 gm bis in 500 ml of glass-distilled water. The solution was de-ionised by adding 10 g of mixed-bed resin and gently stirring for 30 minutes at 4°C. Mixed-bed resin was removed by filtration and the gel stock was degassed for 30 minutes using a vacuum pump. A 5% polyacrylamide gel was prepared by mixing 6 ml of 50% gel stock, 6 ml of 10x TBE, 48 ml water, 400 ul of freshly prepared 25% (w/v) APS and 100 ul

TEMED. Gel dimensions were 20 x 40 x 0.05 cm or 17 x 26 x 0.05 cm. Pre-electrophoresis was at 100V for 20 minutes and electrophoresis at a constant 500 V.

2.18.4(b) Sequencing (Denaturing) Gels

Stock gel solution (6% polyacrylamide / 8 M urea in 1x TBE, acrylamide:bis 19:1) was prepared by dissolving 57 gm acrylamide, 3 gm bis and 480 gm urea in 400 ml glass-distilled water at room temperature. The solution was made to 900 ml with glass-distilled water and de-ionised with 30 g mixed-bed resin for 30 minutes. After removing the mixed-bed resin by filtration, 100 ml of 10x TBE was added and the solution degassed as described in 2.18.4(a).

Polymerisation was carried out by adding 480 ul of freshly prepared 25% (w/v) APS and 120 ul of TEMED to 80 ml of gel stock solution. Gel dimensions were either 20 x 40 x 0.025 cm or 40 x 40 x 0.025 cm. After allowing 30 minutes for polymerisation (warming the plate sandwich to 37^{o}C aided polymerisation), the gel was set up vertically in the apparatus. Pre-electrophoresis was for 30 minutes at 1200 V with the comb in place to prevent well distortion. The comb was removed immediately prior to loading and the wells flushed with 1x TBE to remove any unpolymerised acrylamide. Running buffer for all polyacrylamide gels was 1x TBE and sequencing gels were electrophoresed at a constant 1500 V.

Band distortion due to localised heating near the centre of the gel was eliminated by sandwiching a plastic bag of the same dimensions as the gel between the outside gel plate and another glass plate, with the aid of 0.2 cm perspex spacers and clamps. The plastic bag was filled with 1x TBE, the temperature was monitored with the aid of a plate thermometer and not allowed to exceed 50^{o}C.

2.18.4(c) Sequencing Gels to Resolve Band Compression

Sequencing gels containing 25% formamide (v/v) were prepared by including 25% formamide (v/v) in the gel stock solution.

Polymerisation required the addition of 600 ul freshly prepared 25% APS (w/v in water) and 140 ul of TEMED to 80 ml of gel stock solution. Pre-electrophoresis was for 2 to 3 hours at 500 V. Electrophoresis conditions were as described in 2.18.4(b).

2.18.4(d) Recovery of Nucleic Acid from Polyacrylamide Gel Slices

Nucleic acid was eluted from gel slices by immersing the slice overnight in 400 ul of gel elution buffer (500 mM ammonium acetate, 10 mM Mg acetate, 1 mM EDTA, 0.1% SDS, pH 7.6) at 37°C with constant agitation (Maxam and Gilbert, 1980). The eluate was carefully collected and 2.5 volumes ethanol added. Nucleic acids were allowed to precipitate at -80°C for 30 minutes and collected by centrifugation (Eppendorf centrifuge, 20 minutes, 4°C). The pellet was washed in 70% ethanol/TE (v:v), rinsed in redistilled ethanol, and dried in vacuo for 10 minutes. Finally, the pellet was dissolved in TE and stored at 4°C.

2.18.4(e) De-ionisation of Solutions

Solutions were de-ionised by adding 10% (w/v) of mixed-bed resin (2.6) and gently stirring the solution at 4°C for 30 minutes. Mixed-bed resin was removed by filtration.

2.18.4(f) Preparation of Dialysis Tubing

Dialysis tubing was boiled in the buffer to be used for 5 minutes and thoroughly washed with the same buffer at room temperature before use. All dialysis tubing was prepared and used on the same day.

2.18.5 AUTORADIOGRAPHY

Fuji Rx medical X-ray film was used for autoradiography and in general, gels were wrapped in plastic (Gladwrap) and exposed at room temperature for at least 15 minutes and up to 24 hours. Longer periods of exposure were at -80°C with Tungstate intensifying screens. All sequencing gels were exposed overnight at room temperature.

Specific conditions for autoradiography of gels are given in the figure legends.

2.19 M13 CLONING

Techniques in this section and the one following (2.20) were based on the procedures of Messing (1983 and personal communication).

2.19.1 Isolation of the M13 Vector DNA

A 20 ml culture of JM101 was grown overnight in M13 minimal medium, subcultured into 800 ml 2x YT broth and grown with aeration to an A_{600} = 0.8. At the same time a 5 ml culture of JM101 was grown to an A_{600} = 0.6 and 200 ul of the required M13 single-strand stock (2.19.6) was added. Incubation proceeded for one hour at 37°C after which the infected 5 ml culture was added to the 800 ml culture and grown for a further six to eight hours. Cells were pelleted by centrifugation (5,000 g, 10 minutes, 4°C) and M13 phage DNA isolated by the plasmid preparation procedure described previously (2.17.2), followed by CsCl equilibrium density gradient centrifugation (2.15.2) to purify the vector DNA.

2.19.2 Preparation of M13 Vector DNA for Cloning

M13 cloning vectors were prepared by digesting the vector DNA (2.19.1) with the appropriate restriction enzyme(s) (2.19.3) and then the mix was made to 0.1 M Tris-HCl, pH 8.0, 0.1% SDS in a final volume of 60 ul. Calf intestinal phosphatase (10-20 units) was added and the reaction allowed to proceed for two hours at 37°C. The solution was diluted to 100 ul with TE, extracted with TE-equilibrated phenol and the DNA precipitated with ethanol as described previously (2.16.3). M13 vector DNA was further purified from any undigested DNA by fractionation on an agarose gel and recovery of the M13 vector DNA from the gel slice (2.18.3). Vector DNA was finally dissolved in TE at a concentration of 20 ng/ul

and stored at 4°C.

2.19.3 Restriction Analysis

Digestion of DNA with restriction endonucleases was carried out for one to four hours at 37°C (or at 65°C for TaqI digestions) with a two to five fold excess of enzyme (2 to 5 units per ug of DNA) in a volume of 20 ul. Digestion buffers were those specified in the manufacturers catalogue, and were stored at -15°C as 10x stocks. For double digestions, the enzyme with the lowest NaCl concentration was used first, the concentration of NaCl was then raised to the appropriate amount and the second enzyme added. Preparative digests of 20 to 50 ug of DNA were carried out in a 50 ul reaction volume and incubated overnight at 37°C. Reactions were terminated by heating to 70°C for 10 minutes and digests were checked for completion by electrophoresis on minigels (2.18.1).

2.19.4 End-labelling and End-filling

DNA restriction fragments to be used for radioactive size markers were end-labelled in a reaction mix containing 50 mM NaCl, 6 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 unit of E. coli DNA polymerase I (Klenow fragment) and 2-4 uM d[α-^{32}P]CTP or d[α-^{32}P]ATP. After incubation at 37°C for 30 minutes, the reaction was terminated by heating to 70°C for 10 minutes.

One tenth of any solution containing DNA restriction fragments to be isolated from polyacrylamide gels (2.18.4(a)(d)) and then cloned (2.19.5), was end-labelled to allow detection of the fragments by autoradiography.

End-filling to create blunt-ended DNA restriction fragments for cloning was carried out in the same reaction as described above except that the labelled nucleotide was replaced with 2 ul of dNTP solution (0.25 mM of each of the four dNTPs in 5 mM Tris-HCl, pH 8.0, 1 mM EDTA).

2.19.5 Ligation and Transfection

The DNA fragment to be cloned was mixed with M13 vector in a donor to vector molar ratio of 3:1. M13 vector DNA digested with the appropriate restriction enzyme(s) was kept at a constant 20 ng per ligation and the required ratio was achieved by adding 10 ng/kb of donor DNA.

Ligation was carried out in a total volume of 20 ul containing 10 mM $MgCl_2$, 0.1 mM ATP, 50 mM Tris-HCl, pH 7.5, 0.1 units (sticky-end ligations) to 0.5 units (blunt-end ligations) of T4 DNA ligase and DNA to be ligated. Tubes were incubated for a minimum of 5 hours (sticky-end ligations) or overnight (blunt-end ligations) at $14^{o}C$.

Competent cells were prepared by inoculating 2x YT broth with a hundred fold dilution of a stationary phase JM101 (grown overnight in M13 minimal medium at $37^{o}C$) and growing the cells with aeration to an A_{600} = 1.0. After chilling on ice for 10 minutes, the cells were pelleted (5,000 g, 5 minutes, $4^{o}C$), resuspended to half the original volume in ice-cold 50 mM $CaCl_2$, and left on ice for at least two hours. Improved transfection frequencies were obtained if the cells were resuspended to one fifth the original volume in ice-cold 0.1 M $CaCl_2$. Competent cells were always prepared and used on the same day.

An aliquot (1-5 ul) of the ligation mix was added to 0.2 ml competent cells in chilled, sterile glass test tubes and kept on ice for 40 minutes. Cells were then heat shocked at $45^{o}C$ for two minutes and 3 ml of molten YT soft agar containing 20 ul of IPTG (24 mg/ml), 20 ul of BCIG (20 mg/ml) and 0.1 ml of log phase (A_{600} = 1.0) JM101 was added and the mixture poured onto an M13 minimal plate. After setting, the plates were incubated

overnight at 37°C and possible M13 recombinants were scored as white plaques, whilst parental M13 phage plaques were blue. The following controls were included with each transfection ; (1) digested and phosphatased vector at 10 ng/plate, (2) digested, phosphatased and religated vector at 10 ng/plate, (3) uncut vector at 1 ng/plate.

2.19.6 Preparation of M13 Phage Stocks

A fresh overnight stationary culture of JM101, grown in M13 minimal medium, was diluted fifty-fold into 2x YT broth and 2 ml dispensed into screw-capped polycarbonate tubes. Cultures were infected with M13 phage by toothpicking white plaques and incubating at 37°C with constant agitation for 5 to 7 hours. One ml of cells was transferred to an Eppendorf tube. All the following centrifugation steps (2.19.6, 2.19.7) were carried out in an Eppendorf centrifuge. After pelleting the cells for 10 minutes at room temperature, the supernatant was carefully transferred into a fresh Eppendorf tube and centrifuged for a further 10 minutes. To each 1 ml of supernatant was added 270 ul of PEG solution (20% PEG w/v, 2.5 M NaCl) and phage particles were allowed to precipitate for 15 minutes at room temperature. The phage pellet was collected by centrifugation at room temperature for 5 minutes and the supernatant withdrawn by aspiration. Tubes were centrifuged again for 2 seconds and any traces of the supernatant were removed. The phage pellet in each tube was finally resuspended in 200 ul of TE. Phage stocks prepared in this manner were kept at 4°C for no longer than one week, or for longer periods at -15°C.

2.19.7 Preparation of Template DNA for Sequencing

Phage stocks (2.19.6) were extracted with one half volume TE-equilibrated phenol (2.16.1) at room temperature, and the phases separated by centrifugation for 3 minutes. The aqueous phase was carefully withdrawn (150 ul), avoiding the interface of the two

phases and 6 ul of 3 M Na acetate, pH 5.2 added, followed by 400 ul of ethanol. Precipitation was allowed to proceed at -15°C overnight or at -80°C for 30 minutes. DNA was pelleted by centrifugation (20 minutes, 4°C), rinsed in 1 ml of 95% ethanol, dried *in vacuo* and finally dissolved in 25 ul of TE.

2.19.8 Identification of Recombinants

2.19.8(a) Direct Gel Electrophoresis

Recombinant phage were identified by taking 2 ul of single-strand phage stock (2.19.6), adding 9 ul water, 1 ul of SDS loading buffer (0.2% bromophenol blue, 0.2 M EDTA, pH 8.3, 50% glycerol, 2% SDS), and heating the solution to 60°C for one hour. DNA liberated from the phage in this manner was electrophoresed on minigels (2.18.1). Recombinant M13 DNA had a slower mobility on these gels than parental M13 DNA containing no insert.

2.19.8(b) Quick Isolation of M13 DNA

To determine the size of the cloned DNA, the analytical plasmid preparation procedure (2.17.1) was used on 1 ml of an infected culture (2.12) to give sufficient RF form for restriction digestion (2.19.3) and analysis on minigels (2.18.1).

2.19.8(c) Complementarity Test for M13 Clones

Two ul of single-strand template DNA (2.19.7) to be tested, was added to 2 ul of reference DNA (MEC18(R) or MEC18(L); Section 2.3) in an Eppendorf tube which also contained 4 ul of 0.1 M Tris-HCl, pH 7.4, 0.1 M $MgCl_2$, 0.5 M NaCl and 2 ul of loading buffer (25% glycerol, 0.1 M EDTA, 0.2% bromophenol blue). After incubation at 60°C for one hour, the samples were electrophoresed on minigels (2.18.1) and the DNA visualised by EtBr staining (0.0004% EtBr in 1x TAE). Clones complementary to the reference clone were able to form hybrid molecules, thereby increasing their molecular weight and retarding their mobility on the gel.

2.20 DNA SEQUENCING

The dideoxynucleotide chain termination sequencing technique (Sanger et al., 1977; Sanger et al., 1980; Schreier and Cortese, 1979) was used with the modifications recommended by A.V. Sivaprasad (1984).

2.20.1 Annealing

The original annealing procedure was to place an Eppendorf tube containing 1 ul (2.5 ng) of M13 universal primer (17-mer; 5'-GTAAAACGACGGCCAGT-3') 6 ul of template DNA (2.19.7), 1 ul 10x TM and 2 ul water, into a boiling water bath for 3 minutes and then allow the tubes to cool to room temperature. An alternative procedure was to place the tubes into a 60^{o}C oven for one hour and then allow the tubes to cool to room temperature.

2.20.2 Polymerisation and Gel Electrophoresis

Radioactive label (8 uCi of d[α-^{32}P]CTP) was dried in vacuo, and redissolved in 2 ul of label supplement (16 uM dCTP in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) and finally the annealed DNA (2.20.1) was added and thoroughly mixed. Reaction mixes were prepared in four Eppendorf tubes by mixing equal volumes of ddNTPs (prepared in TE) and dNTPs (prepared in 5 mM Tris-HCl pH 8.0, 0.1 mM EDTA) and dispensing 2 ul into each of the four tubes. The DNA-label solution (2 ul) was added to each tube and finally, 2 ul of Klenow enzyme solution (5 units of Klenow enzyme diluted to 0.5 units/ul in 1x TM just before use) was dispensed onto the side wall of the reaction tubes. Final concentrations of ddNTPs and dNTPs were as follows:

A:	2 uM dATP,	35 uM dGTP,	35 uM dTTP,	80 uM ddATP
C:	25 uM dATP,	25 uM dGTP,	25 uM dTTP,	15 uM ddCTP
G:	35 uM dATP,	2.5 uM dGTP,	35 uM dTTP,	50 uM ddGTP
T:	35 uM dATP,	35 uM dGTP,	2.5 uM dTTP,	130 uM ddTTP

Sequencing reactions were commenced by a 2 second centrifugation to mix the enzyme solution with the reaction mix, and were incubated at 37°C for 15 minutes. Reactions were "chased" by adding 2 ul of dNTP-enzyme solution (0.25 mM of each dNTP and 0.025 units/ul of Klenow enzyme solution in 1x TE, prepared immediately before use) to each tube and incubating for a further 15 minutes at 37°C. The reaction was terminated by adding 4 ul of formamide loading buffer (95% de-ionised formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10 mM EDTA) and the tubes were heated to 100°C for 3 minutes. Samples (0.5 ul) were loaded onto sequencing gels and electrophoresed (2.18.4(b)).

After electrophoresis one of the glass plates was removed and the gel fixed in 10% acetic acid. Addition of 20% ethanol prevented the gel from swelling and wrinkling. The gel was dried in a 110°C oven for 15 to 45 minutes and then autoradiographed (2.18.5).

2.20.3 Sequencing with Radioactive DNA Restriction Fragments as Primers

Radioactive primers for DNA sequencing were prepared by primer extension of M13 clones, digestion with the appropriate restriction enzymes (see figure legends) (2.21) followed by elution of the radioactive fragment from polyacrylamide gels (2.18.4(d)).

Annealing conditions for these radioactive primers to the M13 single-strand template DNA were 100°C for 3 minutes followed by incubation at 60°C for one hour and slow cooling to room temperature. After annealing, 8 uM dCTP was added to the DNA-label solution (2.20.2) to compensate for the ommision of d[α-^{32}P]CTP. The remainder of the procedure was exactly as described previously (2.20.2).

2.21 PREPARATION OF RADIOACTIVE DNA PROBES

M13 single-strand clones with inserts of the same polarity as the mRNA were used to prepare ^{32}P-DNA probes where the strand complementary to the mRNA was made radioactive. This procedure was adapted from the method of Bruening et al. (1982).

M13 17-mer (2.5 ng) universal primer (5'GTAAAACGACGGCCAGT-3') was annealed to 8 ul of M13 clone DNA (2.19.7) in 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$ at 60^{o}C for one hour and then the solution was allowed to cool to room temperature. Primer extension was carried out using the Klenow enzyme (1 unit), in a final volume of 20 ul in 1x TM. The tube also contained 1.5 uM each of d[α-^{32}P]ATP, d[α-^{32}P]CTP and the other two unlabelled dNTPs (each at 50 uM), for 15 minutes at 37^{o}C. After a 5 minute "chase" with 2 ul of all dNTPs (each at 0.25 mM), the Klenow enzyme was inactivated by heating to 70^{o}C for 10 minutes. The extended product was digested with the restriction enzyme(s) described in the figure legends, for 4 hours at 37^{o}C (2.19.3), and the resulting fragments fractionated by electrophoresis on a 5% polyacrylamide gel (20 x 40 x 0.05 cm) (2.18.4(a)).

The radioactive DNA fragment to be used as a probe was located after autoradiography (2.18.5), by comparison of its size to radioactive DNA size markers (2.9.1, 2.19.4). After excising the required gel slice, the DNA was eluted overnight at 37^{o}C (2.18.4(d)) and concentrated by ethanol precipitation (2.16.3).

2.22 RNA ANALYSIS

All procedures for analysis of RNA required care to be taken in avoiding ribonucleases. Gloves were worn at all times, all glassware was sterilised in a 110^{o}C oven overnight and all other equipment autoclaved or immersed in 1 M KOH for 15 minutes and rinsed thoroughly with sterile glass-distilled water.

2.22.1 RNA Preparation

This method was adapted from a procedure by Court et al. (1980) and a protocol supplied by G. Christie (personal communication).

A fresh stationary phase bacterial culture was diluted fifty fold into L broth, incubated with aeration at 30°C to an A_{600} = 0.8 (5x 10^8 cells/ml) and heat induced by transfer to a 39°C water bath. Aliquots of 10 ml were taken at the times indicated in the figure legends, placed into 50 ml polypropylene tubes and then immediately put on ice. NaN_3 was added to a final concentration of 0.02 M to stop cell metabolism, and the aliqouts were kept chilled on ice until all time samples were collected.

Cells were pelleted (3,000 g, 4°C, 10 minutes) and resuspended in 2 ml of freshly prepared 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 10 mM NaN_3, 4 mg/ml lysozyme. Lysis was accomplished by freezing the samples in a dry-ice/ethanol bath and then placing the tubes immediately into a 20°C water bath and finally allowing the lysate to thaw to room temperature. SDS was added to a final concentration of 0.2% and the tubes placed at 45°C for 3 minutes to ensure complete lysis and to inhibit the action of ribonucleases. The freeze/thaw cycle was repeated once more after SDS treatment.

Two extractions with equal volumes of phenol, equilibrated with RNA buffer (20 mM Na acetate, pH 5.2, 20 mM KCl, 10 mM $MgCl_2$), were carried out and nucleic acids precipitated by addition of one tenth volume of 3 M Na acetate pH 5.2 and 2.5 volumes RNAse-free ethanol. After incubation at -80°C for 30 minutes, nucleic acids were pelleted by centrifugation (10,000 g, 4°C, 20 minutes), the pellet rinsed in RNAse-free ethanol, dried in vacuo and finally redissolved in 0.1 mM EDTA, pH 8.0.

2.22.2 Removal of DNA from RNA Preparations

A simple method to remove contaminating DNA (and residual protein) was based on the procedure of Glisin et. al. (1974) and relies on the high buoyant density of RNA, which allows RNA to pellet in CsCl solutions whereas both DNA and protein, which have lower buoyant densities, do not pellet.

The RNA sample in 0.1 mM EDTA, pH 8.0 (2.22.1) was diluted to 4.5 ml in the same buffer and an equal volume of 7.5 M CsCl added with thorough mixing. The solution was carefully overlayed onto a 2.5 ml pad of 5.7 M CsCl in a 10 ml polyallomer tube. After centrifugation (28,000 rpm, 20°C, 16 hours, SW41 rotor), 9 ml of the supernatant was carefully removed by aspiration and the tubes cut below the level of the CsCl pad with a flamed scalpel blade. The remaining supernatant was removed, the gelatinous pellet rinsed with RNAse-free ethanol and redissolved in 0.1 mM EDTA, pH 8.0.

2.22.3 Estimation of RNA Concentration and Storage of RNA

RNA concentrations were determined on a Zeiss PMQ II spectrophotometer by measuring the A_{260} (A_{260} = 1.0 corresponds to a concentration of 40 ug/ml). RNA was stored at -80°C until required. Yields of DNA-free RNA, after centrifugation through the CsCl pad, were generally 1.5 mg/10 ml aliquot of cell culture.

2.22.4 Northern Transfer and Hybridisation

2.22.4(a) Glyoxlyation and Transfer from Agarose Gels

Nucleic acid samples were denatured with 1 M de-ionised glyoxal (2.18.4(e)) in 10 to 20 ul of 10 mM Na phosphate, pH 6.5, 0.1 mM EDTA at 50°C for 45 minutes. DMSO was omitted from the glyoxylation procedure of McMaster and Carmichael (1977) because of excessive breakdown of RNA, even when redistilled DMSO was used (K. Gordon personal communication). The omission of DMSO did not

affect the denaturation process (data not shown).

Samples were electrophoresed on 1.8% agarose gels in 10 mM Na phosphate pH 6.5, at 30 mA for 3 hours (2.18.2). Non-radioactive RNA marker tracks were detected by staining with EtBr (0.0004% ug/ml in 10 mM Na phosphate pH 6.5) and photographed under shortwave UV-light. Radioactive nucleic acid markers were either transferred to nitrocellulose or the track cut from the agarose gel and immediately autoradiographed (see figure legends).

RNA was transferred unidirectionally to nitrocellulose (Schleicher and Schuell, BA85, 0.45 um) by blotting as described by Thomas (1980), using 20x SSC as the transfer buffer. Bidirectional transfers were by the blotting procedure of Smith and Summers (1980), using 20x SSC as the transfer buffer. After transfer, the filters were air-dried for 30 minutes, baked in vacuo at 80°C for two hours and then placed into 400 ml of 10 mM Tris-HCl, pH 8.0, at 100°C. The filter was agitated slowly until the buffer reached room temperature. This procedure was recommended for removal of all residual glyoxal which may interfere with the hybridisation reaction (Thomas, 1983).

2.22.4(b) Hybridisation and Washing

Prehyridisation and hybridisation conditions of ^{32}P-DNA probes to nitrocellulose bound RNA, were as described by Thomas (1980). Probes were all heat denatured at 100°C for 5 minutes and diluted into prechilled hybridisation buffer. After overnight hybridisation at 42°C, the hybridisation buffer was removed and the filters were washed four times for 5 minutes at room temperature in 250 ml 1x SSC, 0.1% SDS, and twice at 60°C in 250 ml 0.1x SSC, 0.1% SDS. Filters were air-dried for 30 minutes, ccvered with plastic film (Vitafilm) and autoradiographed at -80°C (2.18.5).

2.22.5 In vitro Transcription of 186 DNA

Phage 186 DNA purified by CsCl equilibrium density gradient centrifugation (2.15.2) and phenol extraction (2.16.2) was used to prepare radioactive RNA markers by *in vitro* transcription as described by Pritchard and Egan (1985).

DNA (5 ug) was added to a reaction mix containing 25 mM Tris-HCl, pH 8.0, 150 mM KCl, 1.6 mM Na_2EDTA, 1.0 mM dithiothreitol, 0.2 mM each of ATP, CTP, UTP, and 0.02 mM GTP in a volume of 24 ul. This mix was added to an Eppendorf tube containing 2 uCi [α-^{32}P]GTP (dried *in vacuo* for 15 minutes) and 5 ug (0.15 ul) of *E. coli* RNA polymerase was then added. After 5 minutes incubation at 37°C (to allow binding of the RNA polymerase), the synthesis reaction was initiated by addition of $MgCl_2$ to a 10 mM final concentration, and the reaction allowed to proceed for a further 10 minutes at 37°C. Nucleic acids were precipitated by addition of 2.5 volumes of RNAse-free ethanol and stored at -15°C until required. Marker 186 RNA samples were prepared by collecting the ethanol precipitate (10,000 g, 4°C, 10 minutes), rinsing the pellet with RNAse-free ethanol and drying the pellet *in vacuo* for 10 minutes before finally redissolving the pellet in 5 ul 0.1 mM EDTA, pH 8.0. Glyoxylation of these samples was as described in 2.22.4(a).

2.22.6 RNA Sequencing by Primer Extension

Nobrega *et al*. (1983) described a procedure to rapidly detect specific RNA transcripts in crude RNA mixtures. Hybridisation of single-stranded or double-stranded radioactive DNA probes to RNA in solution, followed by gel electrophoresis and autoradiography, allowed the detection of specific DNA/RNA hybrids. This procedure was modified to incorporate the liquid hybridisation procedure (Vogelstein and Gillespie, 1977) optimised for use with phage 186

DNA and RNA by Hooper, (1979). Furthermore, by selecting a probe near the 5'-terminus of the RNA transcript, the hybrid(s) could be isolated from the gel and used to determine the DNA sequence corresponding to the 5'-terminus of the RNA transcript, using the dideoxynucleotide chain termination sequencing technique and AMV reverse transcriptase (2.22.6(b)).

2.22.6(a) Annealing of the Radioactive DNA Primer to in vivo RNA and Isolation of the Primed-RNA

Radioactive DNA primer (2.21) was added to 200 ug of RNA (2.22.2), precipitated with ethanol and redissolved in 20 ul of de-ionised formamide, 66 mM KH_2PO_4, pH 7.0. The mixture was heated to 90°C for 5 minutes, annealed for 4 to 5 hours at 45°C and then fractionated on a 5% polyacrylamide gel (17 x 26 x 0.05 cm) in 1x TBE, using a 10 cm wide well to prevent overloading of the gel.

The RNA/DNA hybrid(s) (primed-RNA) were detected by autoradiography (2.18.5), the gel slice excised from the gel, cut into 10 x 1 cm slices and the primed-RNA eluted (2.18.4(d)). Pooled eluates were extracted with phenol (2.16.3), precipitated with RNAse-free ethanol and redissolved in 10 to 15 ul of 0.1 mM EDTA, pH 8.0.

2.22.6(b) Sequencing With AMV Reverse Transcriptase

Primed-RNA (2 ul, 2.22.6(a)) was aliquoted into each of four tubes and to each tube was added 2 ul of the following dNTP/ddNTP mixes in 1x AMV reverse transcriptase buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT)

A: 0.02 mM dATP, 0.2 mM dCTP, dGTP, dTTP, 0.01 mM ddATP

C: 0.02 mM dCTP, 0.2 mM dATP, dGTP, dTTP, 0.002 mM ddCTP

G: 0.02 mM dGTP, 0.2 mM dATP, dCTP, dTTP, 0.002 mM ddGTP

T: 0.02 mM dTTP, 0.2 mM dATP, dCTP, dGTP, 0.01 mM ddTTP

AMV reverse transcriptase (2 units, in 1x reverse transcriptase

buffer) was added to give a final volume of 5 ul and the tubes incubated at 37°C for 30 minutes. An equal volume of formamide loading buffer (2.20.2) was then added to each tube. The tubes were heated at 100°C for 3 minutes, chilled immediately on ice and electrophoresed on a sequencing gel (2.18.4(b)). After electrophoresis, the gel was fixed in 20% ethanol/10% acetic acid (v/v in water), dried at 110°C for 15 minutes and autoradiographed at -80°C with an intensifying screen.

2.22.7 Determination of 5'-ends of RNA Transcripts by Primer Extension

This procedure was based on the method described by McKnight et al. (1981) with modifications recommended by R. Sturm (personal communication).

Radioactive DNA restriction fragments to be used as primers were prepared as described in Section 2.21.

Radioactive DNA primer and 5 ug of in vivo mRNA (2.22.2) were precipitated with ethanol (2.16.3), redissolved in 10 ul 200 mM NaCl, 10 mM Tris-HCl, pH 8.3, and heated at 100°C for 3 minutes. After annealing at 60°C for 90 minutes, the tubes were allowed to cool to room temperature and 24 ul of reaction mix added to give a final concentration of 10 mM Tris-HCl, pH 8.3, 10 mM $MgCl_2$, 10 mM dithiothreitol, 500 uM of each of the four dNTPs, and 60 mM NaCl. One ul (15 units) of AMV reverse transcriptase was added and the extension reaction was allowed to proceed for one hour at 42°C. Nucleic acids were precipitated with RNAse-free ethanol, the pellet rinsed in ethanol, dried in vacuo and redissolved in 2 ul 0.1 mM EDTA. An equal volume of formamide loading buffer (2.20.2) was added and the samples were heated, electrophoresed and autoradiographed as described in Section 2.20.2.

2.23 COMPUTER ANALYSIS

DNA sequence data was stored and manipulated by the data base management system of Staden (1980).

2.23.1 Predicting Protein-Coding Frames (GENE)

A method of assessing DNA sequences for their protein-coding potential, based on codon usage (Staden and McLachlan, 1982), was used. Program GENE, was used for analysis of protein-coding potential in phage 186 and employed the codon usage of *E. coli* (Chen *et al.*, 1982) genes or the lambda *cI* gene (Daniels *et al.*, 1982) as standards. Codon usage of all possible reading frames was compared (in window lengths of 30-40 codons) to the standard, and the frame which had a similar codon usage pattern to the standard was predicted to be a protein-coding frame. Calculation of the score was described in detail in Staden and McLachlan (1982).

2.23.2 Searching for Signals (SCAN)

Program SCAN predicts the location of potential transcription promoters using a weight matrix to evaluate each section of the DNA sequence by the same principal used in the program of Staden (1984c). Promoters were predicted using a weight matrix composed of the frequency of each base at each position of the 112 *E. coli* promoters compiled by Hawley and McClure (1983). Variable spacings (15-21 bp) between the -10 and -35 regions were accommodated by using two matrix blocks of 25 bases (around the -35 region) and 19 positions (around the -10 region), with spacings of 0 to 6 bases between the two blocks. Scores were obtained by addition of logarithms (base 10) of the raw base frequencies. Arbitrary threshold scores were determined, above which most of the promoters of Hawley and McClure, (1983) scored (77 out of 112), and above which few random sequences (approximately 1 in 2000 kb) scored. The thresholds used

for various spacings between the -10 and -35 regions were 66.0 (17 bp), 66.8 (16 and 18 bp), 67.8 (15 and 19 bp), 68.3 (14 and 20 bp) and 68.8 (21 bp). Final scores were expressed as scores minus threshold. Mulligan et al. (1984) have shown a good correlation between a kinetic parameter related to promoter strength in vitro (log $K_B k_2$) and a promoter homology score calculated from the sequence using a method very similar to SCAN. Promoter scores from SCAN also correlate well with log $K_B k_2$ (I. Dodd, personal communication) and therefore show some relation to promoter strength. The SCAN program was also modified to allow weight matrix analysis of protein sequences, for the purpose of searching for potential DNA-binding regions in the 186 proteins. The weight matrix was composed of the number of occurrences of each amino acid at each position in the alignment of 21 DNA-binding proteins by Pabo and Sauer (1984). Scores were obtained by addition of these raw frequencies.

GENE and SCAN programs were written by I.B. Dodd and all computer analysis was performed on a DEC PDP-11 minicomputer.

SECTION 3

DETERMINATION OF THE DNA SEQUENCE

3. DETERMINATION OF THE DNA SEQUENCE

3.1 INTRODUCTION

An essential prerequisite to studying the control of gene expression is a knowledge of the exact gene content of the region of interest. Genetic studies (Hocking and Egan, 1982a; Younghusband et al, 1975) have identified three genes (*int*, *cI*, *B*) in the early control region, however the potential coding capacity of the DNA in this region indicates more genes will be found (Section 1.3.1). The most direct approach to locate these genes is to determine the DNA sequence of the region and identify the potential genes. DNA signals important in the control of gene transcription can also be searched for.

A 3.3 kb *Pst*I(65.5%-76.3%) fragment of 186 DNA, conveniently spans the region where several early genes (*B*, *cI* and *int*) have been mapped (Hocking and Egan, 1982a; Finnegan and Egan, 1979) and was therefore the fragment of choice for sequencing (see also Section 1.8). The strategy chosen for sequencing was to subdigest the 3.3 kb *Pst*I(65.5-76.3%) fragment with restriction enzymes that recognise four base-pair sites, isolate the various fragments by fractionation on 5% polyacrylamide gels, and clone each fragment into M13 vector DNA. A series of overlapping clones generated by this approach would allow the DNA sequence of the 3.3 kb *Pst*I fragment to be determined. The dideoxynucleotide chain termination technique (Sanger *et al*., 1977; Sanger *et al*., 1980; Schreier and Cortese, 1979) was used for DNA sequencing, with modifications recommended by A.V. Sivaprasad (personal communication).

Phage 186*c*I*ts*p and 186*c*I*ts*p DNA will be referred to continually throughout this work and will be denoted simply as phage 186 and 186 DNA respectively.

3.2 RESULTS AND DISCUSSION

PstI digestion of phage 186 DNA generates a large number of PstI restriction fragments (22 PstI cleavage sites, see Fig. 3.1 track 2) and only the fragments from the right third of the 186 genome have been mapped (Saint and Egan, 1979). The largest PstI fragment was sized at 3.3 kb (65.5%-76.3%) (Saint and Egan 1979), but as reported later (see Fig. 4.5), the size determined from the DNA sequence was 3561 bp. This fragment will, for further discussion, be referred to as the 3.6 kb PstI (65.5%-77.4%) fragment.

To confirm the identity of this 3.6 kb fragment, the PstI digested 3.6 kb fragment was further restricted with XhoI. The XhoI restriction site is unique on the 186 genome, mapping at 68% (Saint and Egan, 1979) and should therefore lie within the 3.6 kb PstI (65.5%-77%) fragment. Double-digestion with PstI and XhoI should yield 2.9 kb and 0.6 kb PstI/XhoI fragments (together with all other PstI fragments) and the 3.6 kb fragment should disappear. After complete PstI/XhoI digestion, the 2.9 kb fragment was observed (Fig. 3.1, track 4) whilst the 0.6 kb fragment comigrated with fragments of similar size and was not clearly visible. A 3.6 kb fragment however remained, indicating the presence of a comigrating fragment not detected by the restriction mapping of Saint and Egan (1979). Similar digests with Hy5 DNA (a P2.186 hybrid phage containing the left 0% to 70.6% of the P2 genome and the right 65.4% to 100% of the 186 cItsp genome; Section 2.2.4) (Fig 3.1, tracks 3 and 5), resulted in the disappearance of the 3.6 kb PstI fragment after PstI/XhoI digestion and the generation of the expected 2.9 kb and 0.6 kb PstI/XhoI fragments. The results from the PstI/XhoI digestion of Hy5 DNA indicated the comigrating fragment in the PstI digest of 186 DNA was located in the 0% to 65.4% region of 186 DNA and

Fig. 3.1. Restriction analysis of 186 and Hy5 DNA

186 and Hy5 DNA (400 ng each) was digested with PstI or doubly-digested with PstI and XhoI. Samples were fractionated on a 1% agarose gel in TAE at 75 mA for two hours. The gel was stained with EtBr and photographed. Sizes of fragments (in kb) are given alongside the tracks.

Track 1. λ DNA, digested with HindIII

2. 186 DNA, digested with PstI

3. Hy5 DNA, digested with PstI

4. 186 DNA, doubly-digested with PstI and XhoI

5. Hy5 DNA, doubly-digested with PstI and XhoI

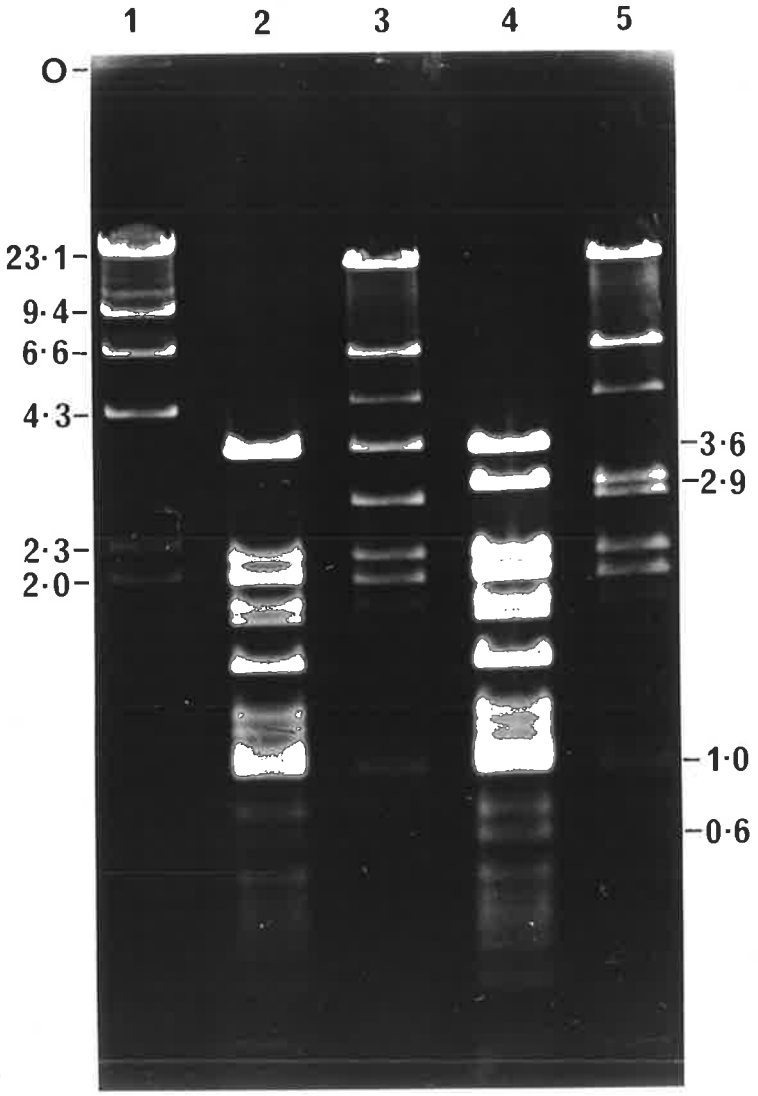
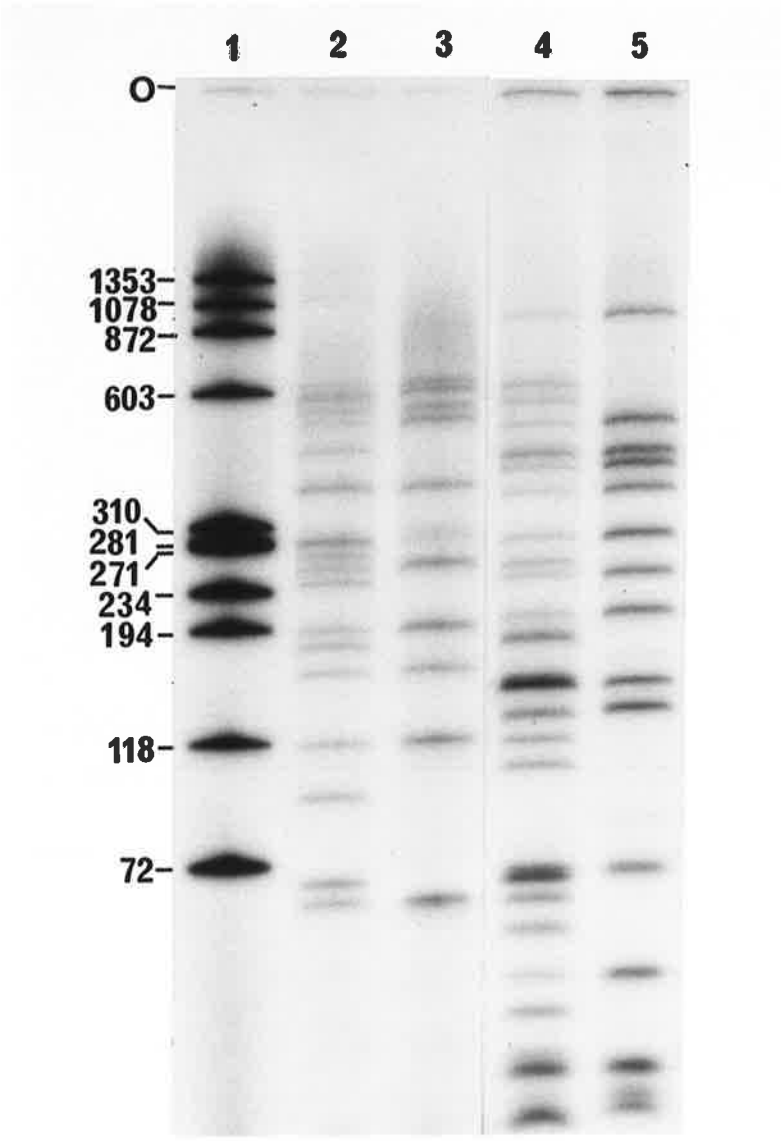

1
2
3
4
5
O
23·1
9·4
6·6
4·3
2·3
2·0
3·6
2·9
1·0
0·6

Fig. 3.2. Sub-digestion of the 3.6 kb PstI Fragment from 186 and Hy5 DNA with HpaII and TaqI

Phage 186 and Hy5 DNA (10 ug each) were digested with PstI and fractionated on 1% agarose gels in TAE. The 3.6 kb PstI fragment from each digest was isolated, sub-digested with HpaII or TaqI and the resulting fragments were end-labelled (Section 2.19.4) and fractionated on a 5% polyacrylamide gel in TBE. Autoradiography was for 4 hours at room temperature. Faint bands appearing in tracks 2 to 5 (>70 bp) represent incomplete digestion products.

Tracks 1. ØX174 DNA, digested with HpaII and end-labelled

2. 3.6 kb PstI fragment from 186 DNA digested with TaqI
3. 3.6 kb PstI fragment from Hy5 DNA digested with TaqI
4. 3.6 kb PstI fragment from 186 DNA digested with HpaII
5. 3.6 kb PstI fragment from Hy5 DNA digested with HpaII


1
2
3
4
5
O
1353
1078
872
603
310
281
271
234
194
118
72

preliminary restriction mapping placed the comigrating fragment at 36% to 48% of the 186 genome (data not shown).

The 3.6 kb PstI fragments from 186cItsp DNA and Hy5 DNA were isolated, sub-digested with HpaII, and fractionated on a 5% polyacrylamide gel (Section 2.19.3, 2.19.4, 2.18.4). A comparison of the two digests allowed the discrimination of HpaII fragments derived from the PstI(65.5-77.4%) fragment of 186cItsp DNA (Fig. 3.2, tracks 4 and 5), which were recovered from the gel and cloned into the AccI site of M13mp7, M13mp8 or M13mp9. TaqI fragments from the 3.6 kb PstI(65.5%-77.4%) fragment of 186cItsp DNA (Fig. 3.2, tracks 2 and 3) were isolated and cloned in the same manner. The DNA sequence generated from these clones revealed the position of several unique restriction sites on the 3.6 kb PstI(65.5%-77.4%) fragment including SacI, SalI, KpnI and XhoI. Restriction fragments generated with various combinations of the above enzymes were identified by size and cloned directly into the appropriate cloning sites of M13mp8 or M13mp9, or were made blunt-ended (Section 2.19.4) and cloned into the SmaI site of M13mp8 or M13mp9.

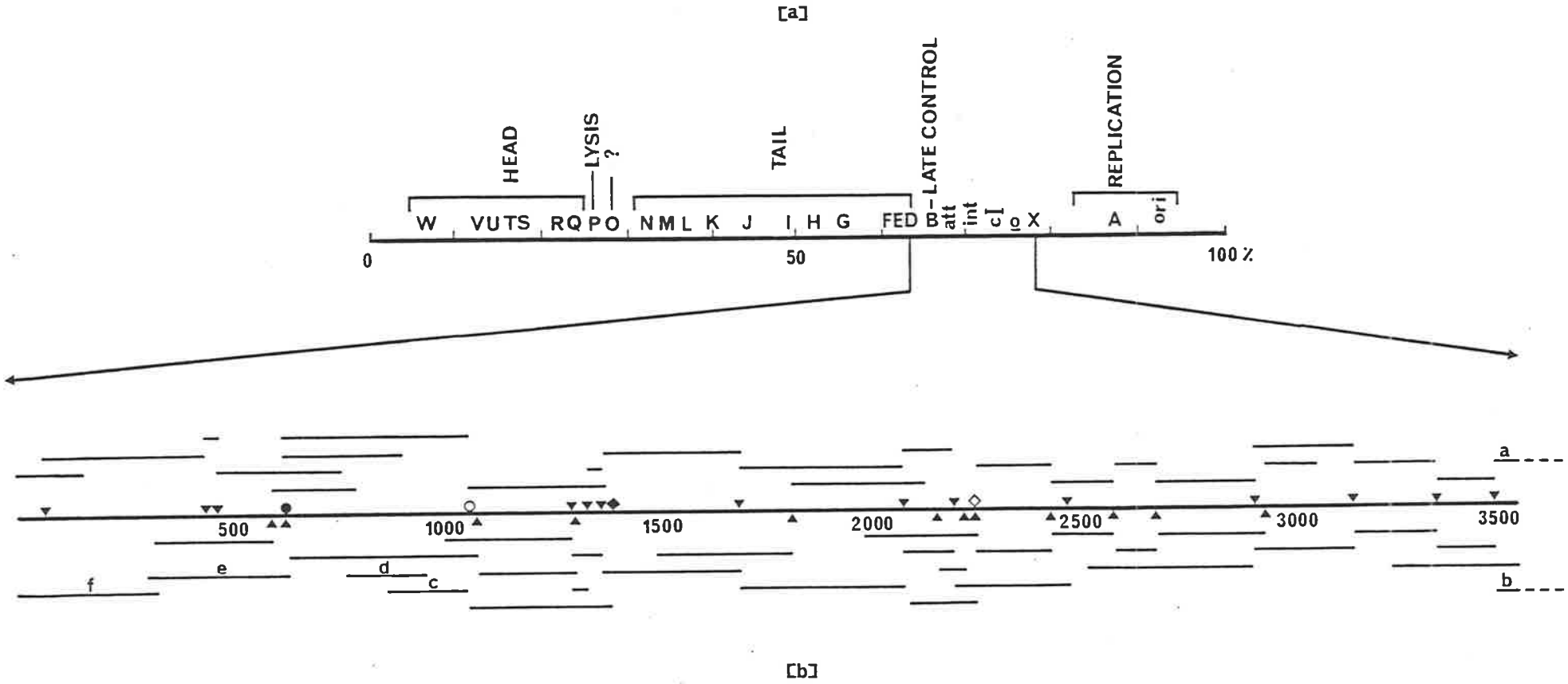
To complete regions of the DNA sequence not determined from both DNA strands, specific radioactive restriction fragments were used as DNA primers. These were synthesized by primer extension of M13 single-stranded clones, followed by digestion with the appropriate restriction enzymes and isolation of the required radioactive restriction fragment (Section 2.21). These specific primers were used to obtain the desired sequence with only minor changes to the sequencing reaction as described in Section 2.20.3.

The region of the 186 genome sequenced is presented in Figure 3.3(a), the sequencing strategy in Figure 3.3(b) and the DNA sequence of the PstI(65.5%-77.4%) fragment in Figure 4.5(a). Both strands of the DNA were sequenced and all restriction sites used to generate

Fig. 3.3. Sequencing strategy for the 3.6 kb PstI (65.5%-77.4%) fragment of 186 DNA

(a) Physical and genetic map of phage 186, and the sequencing strategy for the 3.6 kb PstI(65.5%-77.4%) fragment. Functions of the genes (Hocking and Egan, 1982a) and the physical mapping (Finnegan and Egan, 1979) have been described previously.

(b) The PstI(65.5%-77.4%) region is expanded to show the sequencing strategy. Horizontal bars correspond in length to the actual number of bases read from individual sequences. Bars above the solid line represent gel readings used to generate the l-strand sequence, and those below the line represent gel readings used to generate the r-strand sequence. Restriction sites used during the course of the sequencing were as follows: HpaII (▼); TaqI (▲); SalI (◇); SacI (◆); KpnI (○); XhoI (●). Sequences (a) and (b) were determined by H. Richardson (manuscript in preparation) and sequence (e) by Pritchard and Egan (1985). Sequences (c), (d) and (f) were determined by primer extension of single-stranded M13 clones (Section 2.20.3) using 114 b KpnI-HhaI (positions 1066-1180), 108 b HhaI-KpnI (positions 957-1065) and 100 b HhaI-HhaI (positions 332-432) primers respectively.


[a]
HEAD
LYSIS
?
TAIL
LATE CONTROL
REPLICATION
W VUTS RQ P O NML K J I H G FED B att int cI o X A ori
0
50
100 %
500
1000
1500
2000
2500
3000
3500
a
b
c
d
e
f
[b]

M13 clones were sequenced over to ensure small restriction fragments did not remain undetected. Unless otherwise specified, all sequence positions and co-ordinates refer to those in Figure 4.5.

One section of the DNA sequence, a palindrome at sequence positions 530 to 550, gave rise to compressed DNA bands when sequenced from either DNA strand and the sequence of this region was resolved by including 25% (v/v) de-ionised formamide in the polyacrylamide gel solution (Section 2.18.4(c)).

SECTION 4

ANALYSIS OF THE DNA SEQUENCE

4. ANALYSIS OF THE DNA SEQUENCE

4.1 INTRODUCTION

An initial step in the analysis of a DNA sequence is to locate open reading frames which identify the potential genes. Open reading frames of at least 40 amino acids, commencing with an initiation codon ATG (or GTG) and preceded by a potential ribosome binding site (Stormo et al., 1982) were searched for. Reading frames were compared with codon usage standards, to assess their protein coding potential by the method of Staden and McLachlan (1982) (Section 2.23.1). This method has been used successfully to identify genes of phage λ (Sanger et al., 1982), however the distribution of codons differs in proteins encoded on different transcription units in λ and the success of detection varies depending on the choice of a codon usage standard. Several codon usage standards were initially used to detect potential 186 genes but the two standards which gave consistently high scores for potential 186 genes in the GENE program (Section 2.23.1), were the codon usage standard of E. coli genes (Chen et al., 1982) and the λ cI gene (Daniels et al., 1982). Signals important in RNA transcription (promoters and terminators) were searched for and patterns of repeated DNA sequences (direct and inverted repeats) were located. The DNA sequence spanning the 186 Δ1 deletion was determined since the left boundary of this deletion was proposed to locate the 186 prophage attachment site att (Section 1.3.1(a)).

4.2 RESULTS AND DISCUSSION

4.2.1 Protein Coding Potential (Prediction of Genes)

Seven open reading frames of at least 40 amino acids which represent potential genes were located and given the prefix CP (Computer Protein), followed by the percentage co-ordinate approximating the position of the initiation codon on the genome e.g. CP76.

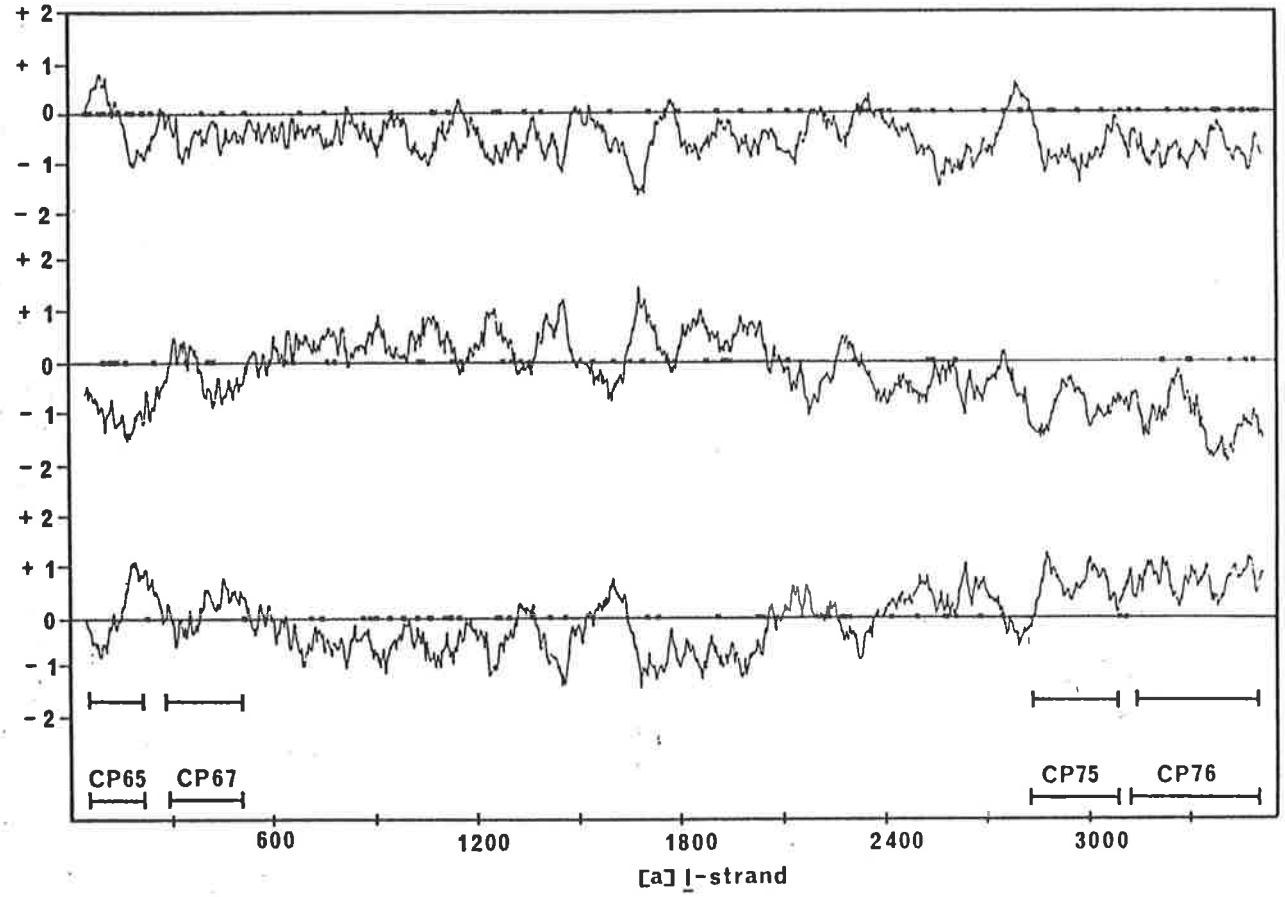
Fig. 4.1. Codon usage plot of the DNA sequence of the 3.6. kb PstI (65.5%-77.4%) fragment

The codon usage plot for the l-strand (a) and the r-strand (b) of the 3.6 kb PstI(65.5%-77.4%) DNA sequence. Along the Y-axis is the sequence position from the 5'-end of the sequence. The position of the predicted coding frame and the name assigned to the frame is given immediatley above the Y-axis.

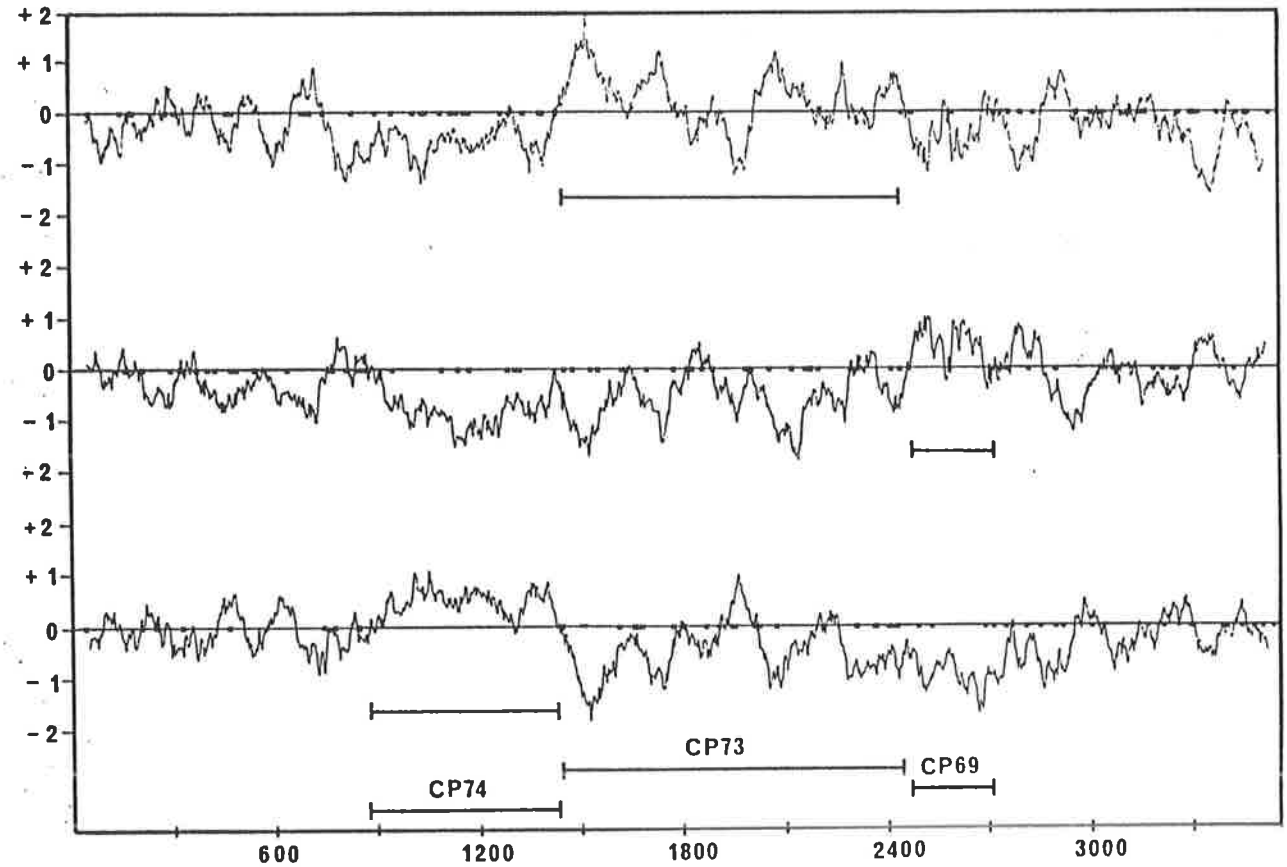
Codon usage plots for each of the three possible reading frames are given above the Y-axis. The score ($\log_{10}(p/1-p)$), which was described in detail by Staden and McLachlan (1982) and in Section 2.23.1, is shown on the X-axis. A positive score for a reading frame indicates that for a section of the DNA sequence (in this case the window was 35 codons), the frame fits the codon usage standard better than the other two frames from the same section. A negative score indicates that the frame shows a poorer fit to the codon usage standard than the other two frames. As a logarithmic score is used, scores of +2 or -2 for example, indicate a frame shows a 100 times better or worse fit to the codon usage standard than the other two frames.

The codon usage standard of E. coli genes (Chen et al., 1982) was used for the l-strand and the codon usage standard of the λ cI gene (Daniels et al., 1982) was used for the r-strand (See Section 4.1).

The position of termination codons in each of the three reading frames is indicated by asterisks. Reading frames predicted to be protein-coding frames are shown immediately below the frame. Additional information used to predict potential protein-coding frames was described in Section 4.1 and Section 5.



[a] l-strand



[b] r-strand

These reading frames are shown by amino acid sequence in Figure 4.5(a) and diagrammatically in Figure 4.5(b).

Four reading frames extending left to right (rightward) with respect to the genetic map were detected by the GENE program (Section 2.23.1) using the codon usage of E.coli genes as a standard (Fig. 4.1). These reading frames extended from sequence positions 1 to 225 (>75aa), 295 to 510 (72aa), 2827 to 3087 (87aa), 3121 to 3561 (>147aa) and were named CP65, CP67, CP75 and CP76 respectively. CP65 and CP76 were only partly represented on the DNA sequence of the PstI(65.5%-77.4%) fragment.

Three reading frames extending right to left (leftward) with respect to the genetic map were detected with the codon usage standard of the lambda cI gene when analysed by the GENE program (Section 2.23.1) (Fig. 4.1). These reading frames extended from sequence positions 2697 to 2122 (192aa), 2119 to 1112 (336aa), 1094 to 846 (83aa) and were named CP74, CP73 and CP69 respectively.

4.2.2 Ribosome Binding Sites

Stormo et al. (1982) analysed the non-random distribution of nucleotides surrounding the initiation codons of known prokaryotic genes and devised a set of rules to locate potential translational initiation sites. The role of RNA secondary structure in the accessibility of the ribosome binding site (RBS) or the initiation codon was not taken into account, nor do these rules predict the translational efficiency of the particular sequence detected. Initially, one of the least stringent rules (rule 2, Stormo et al., 1982) was used and the DNA sequence was scanned. A potential ribosome binding site was located preceding the ATG initiation codon of all predicted protein coding frames except CP75 (reading frame CP65 starts to the left of the PstI (65.5%) site). Potential ribosome binding sites for CP76, CP73 and CP69 were also detected by the

Table 4.1 Ribosome binding sites

Predicted Protein	Ribosome Binding Site (RBS)[a]	Stormo Rule[b]	Sequence Position[c]
CP67(B)	AAGG- 7 -ATG	5	295
CP69	GGAG- 5 -ATG	6	1094
CP73(Int)	GAGG- 7 -ATG	6	2119
CP74(cI)	AGGT- 9 -ATG	4	2697
CP75	GGGA- 10 -ATG[d]	-	2827
CP76	AGGA- 6 -ATG	6	3121

Notes to Table 4.1

a. The numbers between the proposed ribosome binding site and the initiation codon refer to the number of intervening bases.
b. Stormo et al. (1982)
c. Sequence position corresponding to the A residue of the ATG initiation codon.
d. Sequence was not detected by the Stormo rules but shows homology to the Shine-Dalgarno sequence (Shine and Dalgarno, 1975).

more stringent rule 6, whilst the predicted ribosome binding sites of CP67 and CP74 conformed to rule 5 and rule 4 respectively as shown in Table 4.1. The region upstream of the proposed initiation codon for CP75 contained a triplet GGA which has some homology to the Shine-Dalgarno sequence (5'-TAAGGAGGTGA-3'; Shine and Dalgarno, 1974) but the spacing between the triplet and the initiation codon of CP75 was 1 bp more than acceptable for rule 3 (Stormo et al., 1982). CP74 has an alternative, but less frequently observed initiation codon GTG, at sequence positions 2667 to 2665. A potential ribosome binding site, AGGA, was detected 8 bp upstream, at sequence positions 2680 to 2677.

4.2.3 Promoters

Three well conserved DNA sequences (-35, -10 and +1 regions) have been shown to be important in the interaction of E.coli RNA polymerase with DNA (Rosenberg and Court, 1979; Siebenlist et al., 1980). This information has been incorporated into a consensus sequence which can be used to search for potential promoters in a DNA sequence. A more powerful technique of locating potential transcription promoters (and other signals in the DNA sequence) is to assign separate values to each base at each position of the promoter, indicating the relative importance of the base at that position. A weight matrix generated in this manner was used in the SCAN program (Section 2.23.2) to locate potential transcription promoters on the 186 DNA sequence and a promoter homology score (Section 2.23.2) was assigned to each potential promoter.

The three highest scoring promoter sequences are recorded in Table 4.2. Promoter pB was located at sequence position 263 (the position of the highly conserved T residue of the -10 hexamer) and would initiate rightward transcription of CP67 (shown to be the B gene in Section 5.2.2). Two promoters were located between

Table 4.2 Promoter sequences

	Sequence Position[b]	Promoter Score[c]	-35	-10
Consensus[a]			TTGACa	TAtAaT
pB	263	2.4	TTCACA---17---TATCAT	
pR	2740	1.2	TTTACT---17---TATATT	
pL	2817	0.7	TTGCGA---17---CATGAT	

Notes to Table 4.2

a. The consensus sequence for the -10 and -35 regions was from the compilation of known promoters of E. coli RNA polymerase (Hawley and McClure,1983). Lower case letters indicate nucleotides that are less highly conserved.
b. The sequence position of the rightmost conserved T residue in the -10 region (Rosenberg and Court, 1979) is given.
c. Promoter score was determined by the SCAN program (See Section 2.23.2). Promoter score for selected promoters from the lambda DNA sequence (Sanger et al., 1982) was 2.6 for pR, 1.3 for pL and -0.7 for pRM.

sequence positions 2736 and 2821 (Table 4.2) which would initiate transcription in the opposite direction and were named pR and pL for rightward and leftward transcription respectively.

The promoter homology score shows some relation to promoter strength *in vitro* (Mulligan *et al.*, 1984; Section 2.23.2) and on this basis, the order of *in vitro* promoter strength would be pB>pR>pL.

4.2.4 Transcription Terminators

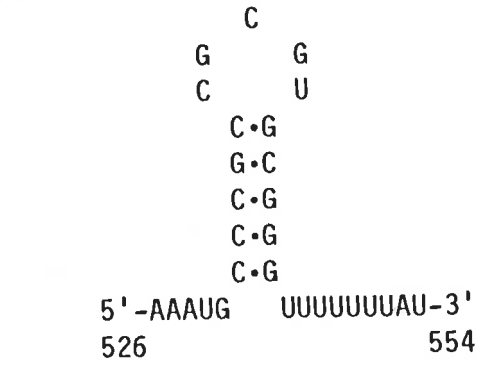
Sites that signal termination of transcription by *E. coli* RNA polymerase in bacterial and phage DNA fall into two major categories; *rho*-dependent and *rho*-independent (Rosenberg and Court, 1979). Terminators that depend on the presence of *rho*-protein for transcription termination to occur (Roberts, 1969), do not show a consistent set of structural homologies and are difficult to predict from the DNA sequence alone. *Rho*-independent terminators are functionally active *in vitro* in the absence of protein factors, other than RNA polymerase, and generally show the following common structural features (Rosenberg and Court, 1979; Holmes *et al.*, 1983); (1) a region of dyad symmetry, rich in GC base pairs, which allows the formation of a stable stem-loop structure in the RNA transcript, (2) a run of consecutive T residues (U residues in the RNA transcript) immediately downstream of the dyad symmetry. These features do not unequivocally characterise the *rho*-independent transcription signal and other DNA sequence features may be involved (Brendel and Trifinov, 1984).

The DNA sequence was scanned using the HAIRPN program (Staden, 1978) to locate DNA sequences which potentially form stem-loop structures (a stem of more than 5 bp with a loop of no more than 20 bases was searched for) and those structures that fulfilled the above criteria were selected. Two stem-loop structures were located at sequence positions 531 to 554 and 643 to 616, which

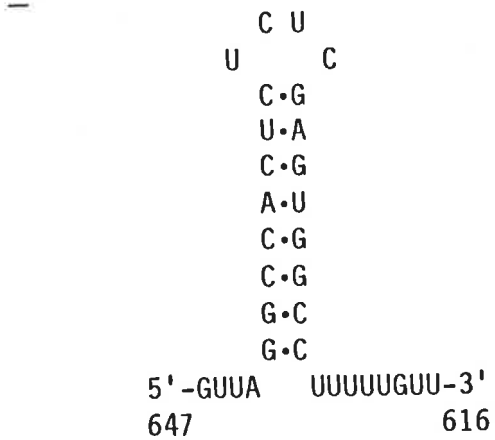
Fig. 4.2. Predicted rho-independent transcription terminators

Terminators tB and tL fulfil the criteria for rho-independent transcription terminators (Rosenberg and Court, 1979) and are presented as the stem-loop structures proposed to form in the mRNA transcript. Below each structure are the DNA sequence positions.

tB


C
G G
C U
C•G
G•C
C•G
C•G
C•G
5'-AAAUG UUUUUUUAU-3'
526 554

tL


C U
U C
C•G
U•A
C•G
A•U
C•G
C•G
G•C
G•C
5'-GUUA UUUUUGUU-3'
647 616

were followed by runs of T residues (7 and 5 T's respectively). The two potential rho-independent transcription terminators were named tB and tL respectively. Terminator tB would terminate rightward transcription from pB, and terminator tL would terminate leftward transcription from pL, assuming in each case the absence of any rho-dependent transcription terminators. The proposed stemloop structures which can form in the RNA transcript are recorded in Figure 4.2.

4.2.5 Direct Repeat DNA Sequences

Repeated DNA sequences R1 to R6 were found between sequence positions 856 to 1081, each 28 bp long with spacings of 39 or 42 residues between the first residue of each repeat (Table 4.3(a)). Three other repeated sequences (OD1, OD2, OD3) were located between sequence positions 2724 and 2810. OD1 and OD3 were each 25 bp long and OD2 was 24 bp long. Repeats OD1 and OD2 had a spacing of 32 bp between the first residue in each repeat, whilst a spacing of 30 bp was found between the first residues of repeats OD2 and OD3 (Table 4.3(b)).

4.2.6 Inverted Repeat DNA Sequences

The largest inverted repeat DNA sequence was located between sequence positions 2714 to 2739 and named OI1, its central palindrome (CAATTG) was repeated at sequence positions 2746 to 2771 (central palindrome of OI2) and 2781 to 2806 (central palindrome of OI3) (Table 4.4). A spacing of 32bp was found between the C residue of the central palindrome of OI1 and OI2 and a spacing of 35bp was found between the same residues of the central palindromes of OI2 and OI3. Other than the central palindromes, OI2 and OI3 do not show the inverted repeat nature of OI1 (Table 4.4). The palindrome CAATTG was also found at sequence positions 2386 to 2391. Many inverted repeat sequences were found between sequence

Table 4.3 Direct repeat DNA sequences

		Spacing[c]	Sequence Position[b]
R6[a]	TCGCCGACATAGGTTCCATCAGGTGCGA	39	856
R5	CCACCAACATAAGTCCCATCTGGGGCAA	39	895
R4	CCAGCAACATAACTTCCATTTGGCGCTA	39	934
R3	CCACCGACATAAGTTCCATCGGGTGCAA	42	973
R2	CCTCCAACATAAGTCCCATTAGGGGCAA	39	1015
R1	CCGTAAACATAGGTACCATCAGGTGCTA		1054
Consensus[d]	CC-cCaACATAaGT-CCATc-GG-GC-A		
OD1	CAATTGGGAGATATATTTTGGCTAA	32	2724
OD2	CAATTG-ATGGCAAGTGTTGGCAAA	30	2756
OD3	CAAATCAATTGCAAACTTTGGCTAA		2791
Consensus[d]	CAAtTg-atggcAaattTTGGCtAA		

Notes to Table 4.3

a. The repeats when translated give an amino acid repeat of consensus $G_5 G_4 V_6 Y_6 T_5 G_6 D_4 P_6 A_6 L_6$

b. Sequence position of the first base of the repeat.

c. Number of bases between the first base of the repeat to the first base of the next repeat.

d. Consensus sequence capital letters represent nucleotides conserved in 6/6, 5/6 or 3/3 of the sequences. Lower case letters represent nucleotides conserved in 4/6 or 2/3 sequences.

Table 4.4 Inverted repeat DNA sequences

		Spacing	Sequence[a] Position
OI1	TACTATCTCTCAA.TTGGGAGATATAT	32[b]	2724
OI2	TAAACCCACGCAA.TTGATGGCAAGTG	35	2756
OI3	AGAGTCAAATCAA.TTGCAAACTTTGG		2791
BI1	ATGATGAATAAAATGTATTCACAAA	32[c]	233
BI2	TTTATGATTATCATTTATTCACGAA		265

Notes to Table 4.4

a. Sequence position of the base referred to in notes b and c.
b. Number of bases between the C residue of the central palindrome CAATTG, to the C of the next palindrome.
c. Number of bases between the first T residue of the conserved sequence TATTCAC in BI1 to that of BI2.

positions 219 to 293 and 530 to 603, corresponding to the regions immediately 5' and 3' to CP67 (data not shown). Two inverted repeats (named BI1 and BI2) were found between sequence positions 219 and 293 which overlapped the -10 and -35 regions of the pB promoter and these are presented in Table 4.4.

4.2.7 att site

Recombinational mapping placed the prophage attachment site (att) between genes B and cI (Hocking and Egan 1982a) but the position of att relative to the int gene was not known.

Hoess and Landy (1978) sequenced several phage λ deletion mutants, isolated by mutagenesis during prophage induction (Parkinson and Huskey, 1971), and found that these mutants had a common endpoint in the core of the λ att site. Deletion mutant Δ1 of 186 was isolated by the same method (Dharmarajah, 1975) and extended from 67.9% to 74.9% (Finnegan and Egan, 1981). The left boundary of this mutant was expected to represent the 186 prophage attachment site (Section 1.3.1(a)) and the DNA sequence of the region spanning the Δ1 deletion was determined in the following manner.

The 3.6 kb PstI (65.5%-77.4) fragment from Hy5 DNA and the 1.7 kb PstI fragment containing the Δ1 deletion from plasmid pEC35 (Section 2.3) were digested with TaqI, end-labelled and fractionated on a 5% polyacrylamide gel. Comparison of the TaqI digests (Fig. 4.3, tracks 2 and 3) revealed a 135 bp TaqI fragment in track 3 which did not co-migrate with corresponding TaqI fragments in the digest of the 3.6 kb fragment, indicating that this TaqI fragment contained the deletion. The 135 bp TaqI fragment was isolated from the polyacrylamide gel, cloned into the AccI site of M13mp7 and sequenced.

The DNA sequence of the 135 bp TaqI fragment is presented in Figure 4.4. Analysis of the DNA sequence revealed that the Δ1

Fig. 4.3. Identification of the 135 bp TaqI fragment spanning the Δ1 Deletion

Five ug of plasmid pEC35 (Section 2.3; Finnegan and Egan, 1981) was digested with PstI and the 1.7 kb PstI fragment containing the Δ1 deletion was isolated after electrophoresis on an agarose gel (Section 2.18.1, 2.18.3). The 3.6 kb PstI(65.5%-77.4%) fragment from Hy5 DNA was also isolated in the same manner.

Both fragments were end-labelled (Section 2.19.4) and fractionated on a non-denaturing 5% polyacrylamide gel in TBE buffer (Section 2.18.4(a)). Autoradiography was for 4 hours at room temperature. The 135 bp TaqI was excised from the gel (Section 2.18.4(d)), cloned into the AccI site of M13mp7 and sequenced (See Figure 4.4).

Tracks 1. ØX174 DNA digested with HaeIII and end-labelled.

2. 3.6kb PstI fragment from Hy5 DNA digested with TaqI.

3. 1.7kb PstI fragment from pEC35 DNA, digested with TaqI.

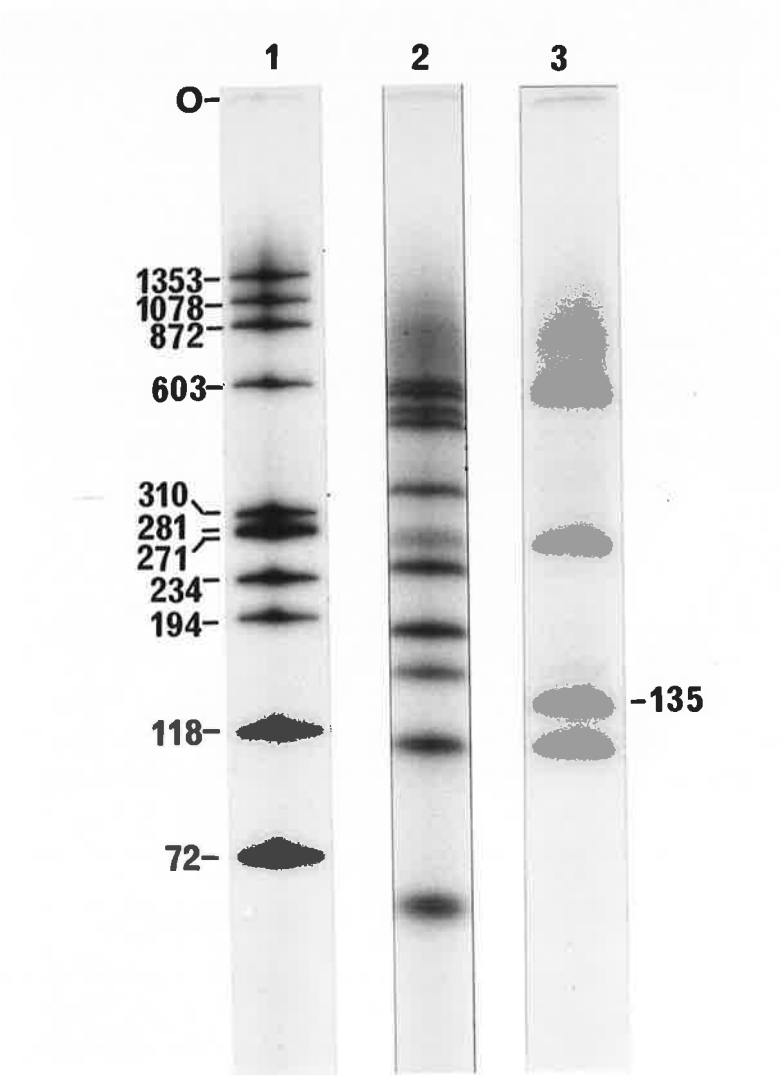
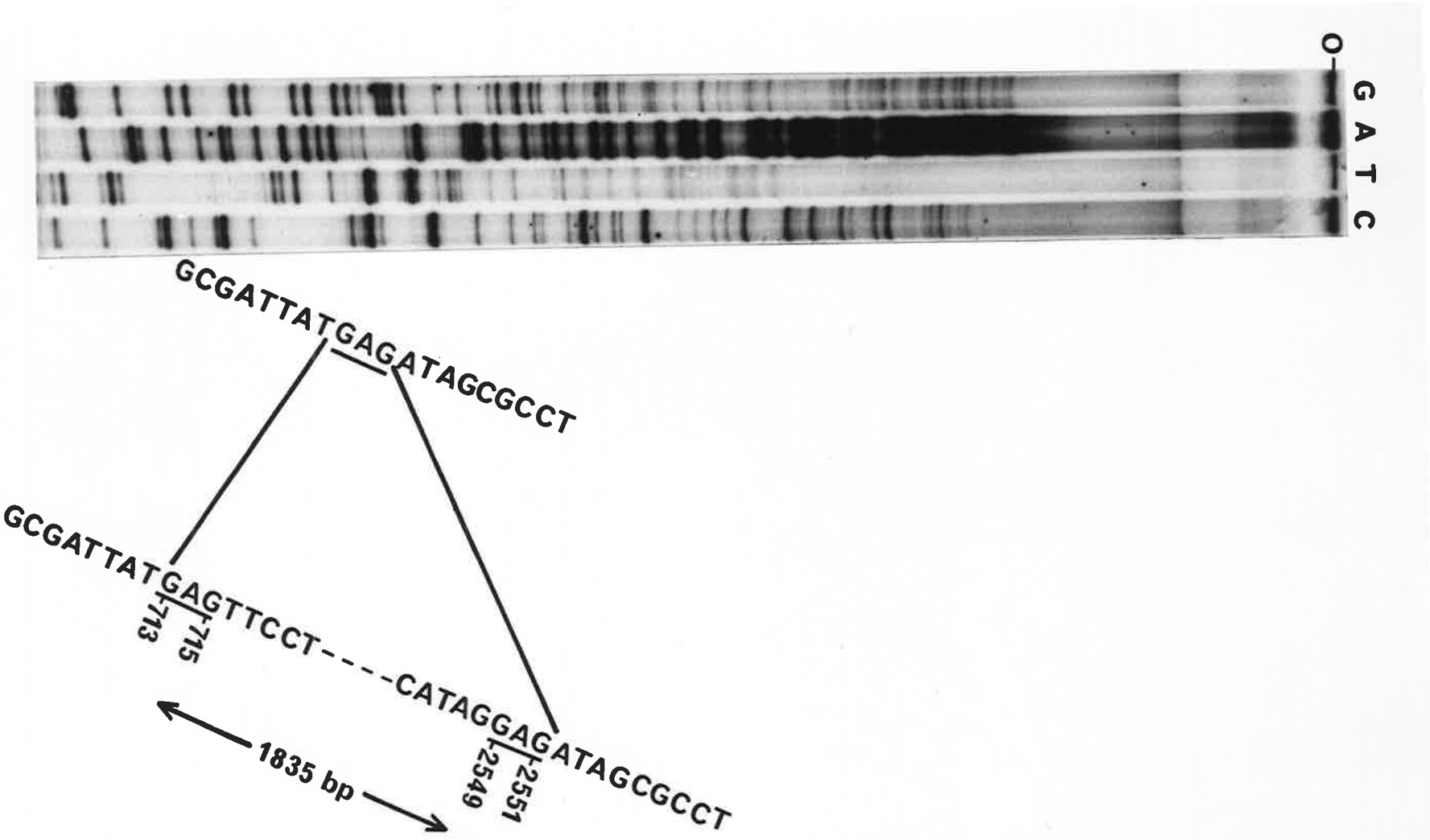

1
2
3
O-
1353-
1078-
872-
603-
310
281
271
234-
194-
118-
72-
-135

Fig. 4.4. DNA sequence of the 135 TaqI fragment spanning the Δ1 Deletion

The 135 bp TaqI fragment predicted to span the Δ1 deletion (Section 4.2.7 and Figure 4.3) was cloned into the AccI site of M13mp7 and sequenced. Although the DNA sequence was determined from both strands, only the DNA sequence of a clone containing the r-strand of the 135 bp TaqI fragment is presented. The deletion occurs beteen two short direct repeats GAG at sequence positions 713 and 715 and GAG at sequence positions 2549 to 2551 as indicated to the right of the autoradiograph. The exact deletion cannot be determined since a GAG triplet is regenerated in the DNA sequence of the 135 bp TaqI fragment. The deletion removes 1835 bp from 67.9% to 74.0%. Electrophoresis and autoradiography conditions were described in Sections 2.18.4(b) and 2.18.5.


O
G
A
T
C
GCGATTATGAGATAGCGCCT
GCGATTATGAGTTCCT----CATAGGAGATAGCGCCT
1713
1715
2549
2551
1835 bp

mutant was a deletion of DNA between the short direct repeats GAG at 713 to 715 and GAG at 2549 to 2551, removing 1835 bp and extending from 67.9% to 74.0%. Deletion Δ1 was originally reported to be from 67.9% to 74.9% (Finnegan and Egan, 1981), based on restriction mapping (Saint, 1979) and heteroduplex analysis (R. O'Connor, unpublished data).

The *att* site was predicted to correspond to the left boundary of the Δ1 deletion and was therefore expected to be located at sequence positions 713 to 715 (67.9%). Recent hybridisation studies have provided evidence that the left end of the Δ1 deletion does not in fact represent the *att* site of phage 186. D. Dodd of this laboratory constructed a series of radioactive DNA probes which cover the 3.6 kb *Pst*I(65.5%-77.4%) fragment and hybridised these probes to lysogenic DNA which had been digested with restriction endonucleases and transferred to nitrocellulose. Using this procedure, the *att* site was mapped in the 70 bp region between terminator *t*B and the *Xho*I site at 67.6% (which is located in the *t*L terminator), some 80 bp at least from the left boundary of the Δ1 deletion (D. Dodd, unpublished results).

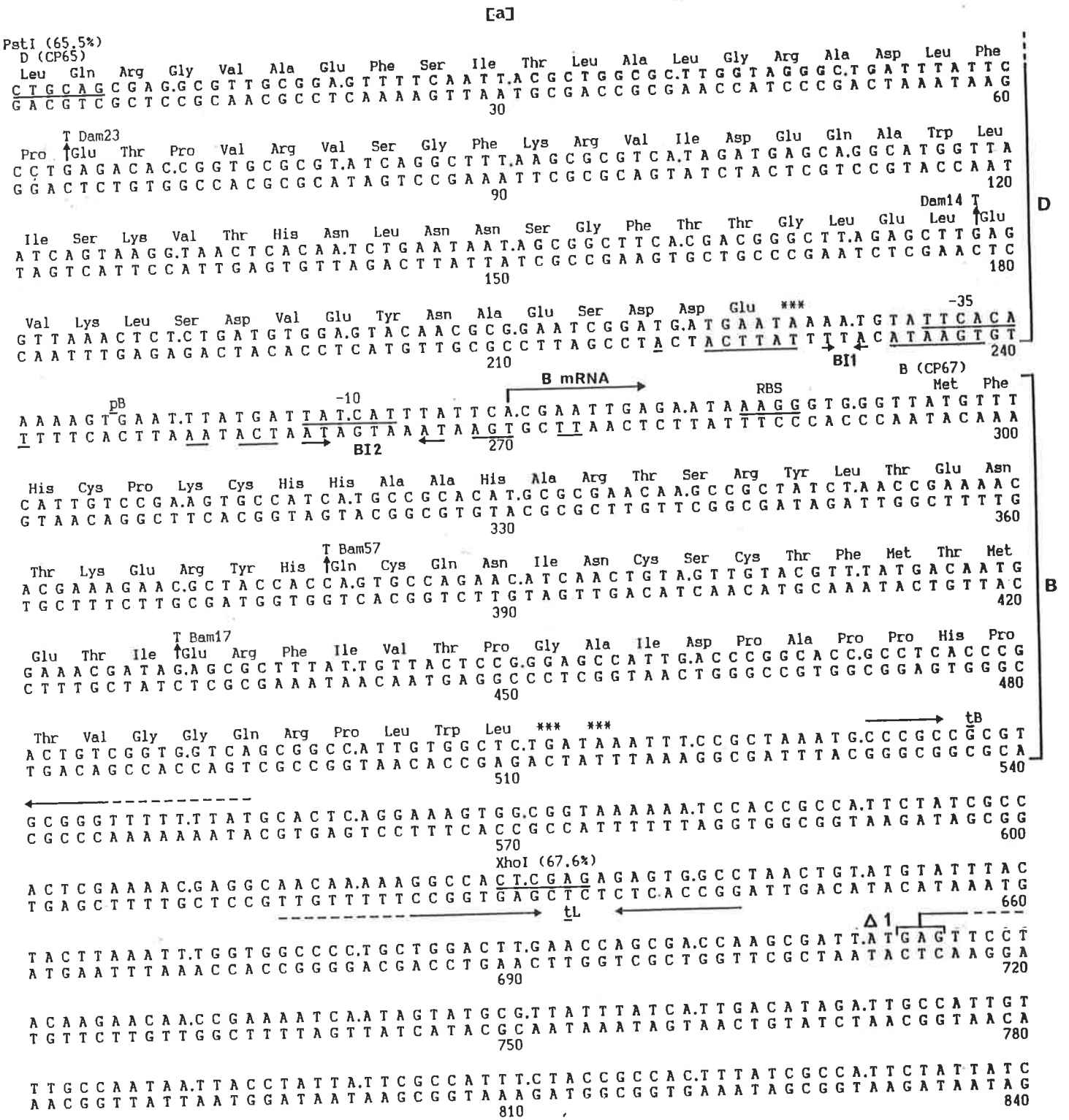
Large deletions (700 b to 2.5 kb) can be generated spontaneously by recombination between short direct repeat sequences (Albertini *et al*., 1982; Jones *et al*., 1982) and it is possible that the Δ1 deletion arose by spontaneous deletion rather than by *int*-mediated recombination.

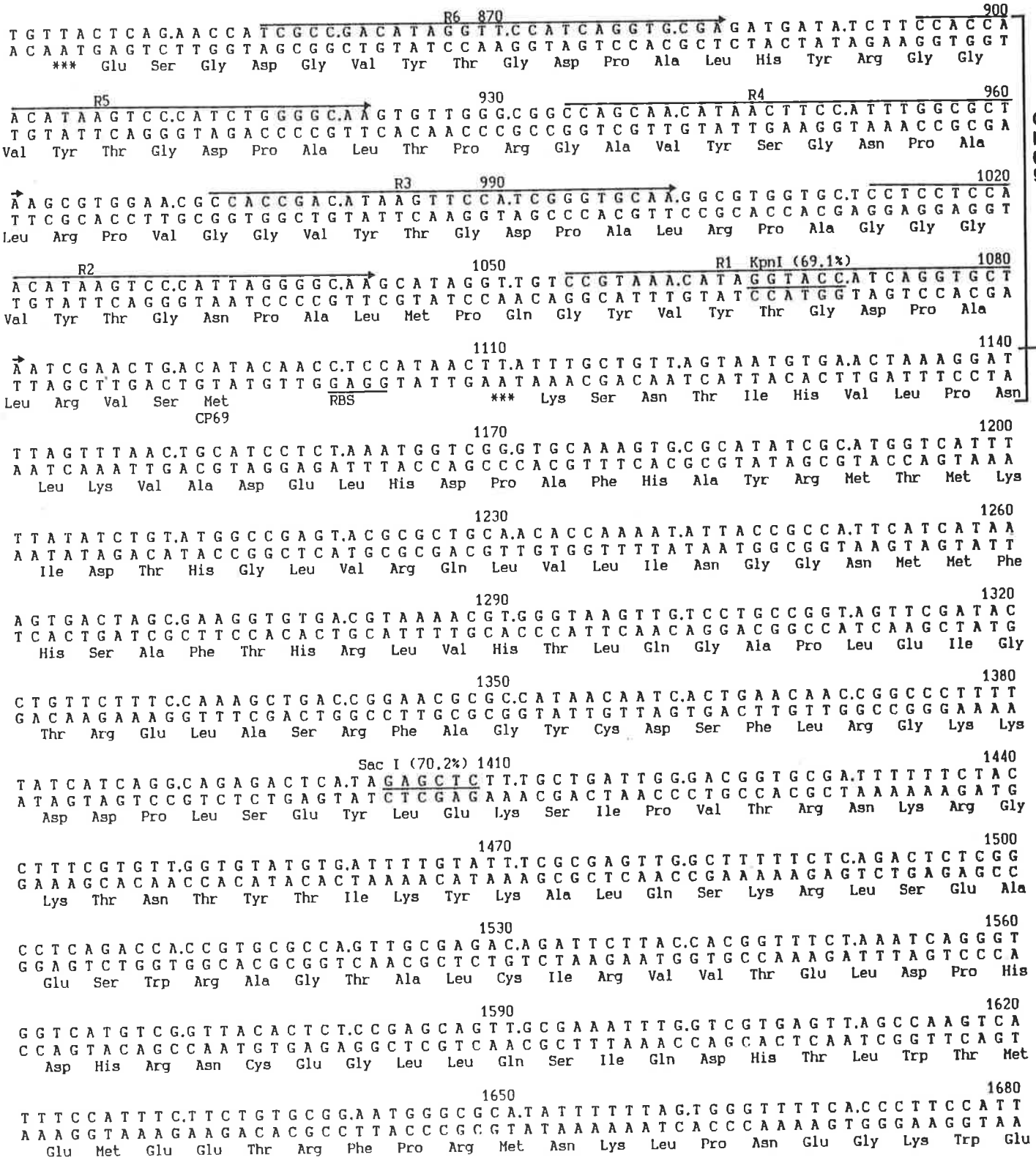
Fig. 4.5. Nucleotide sequence of the 3.6 kb PstI(65.5%-77.4%) restriction fragment

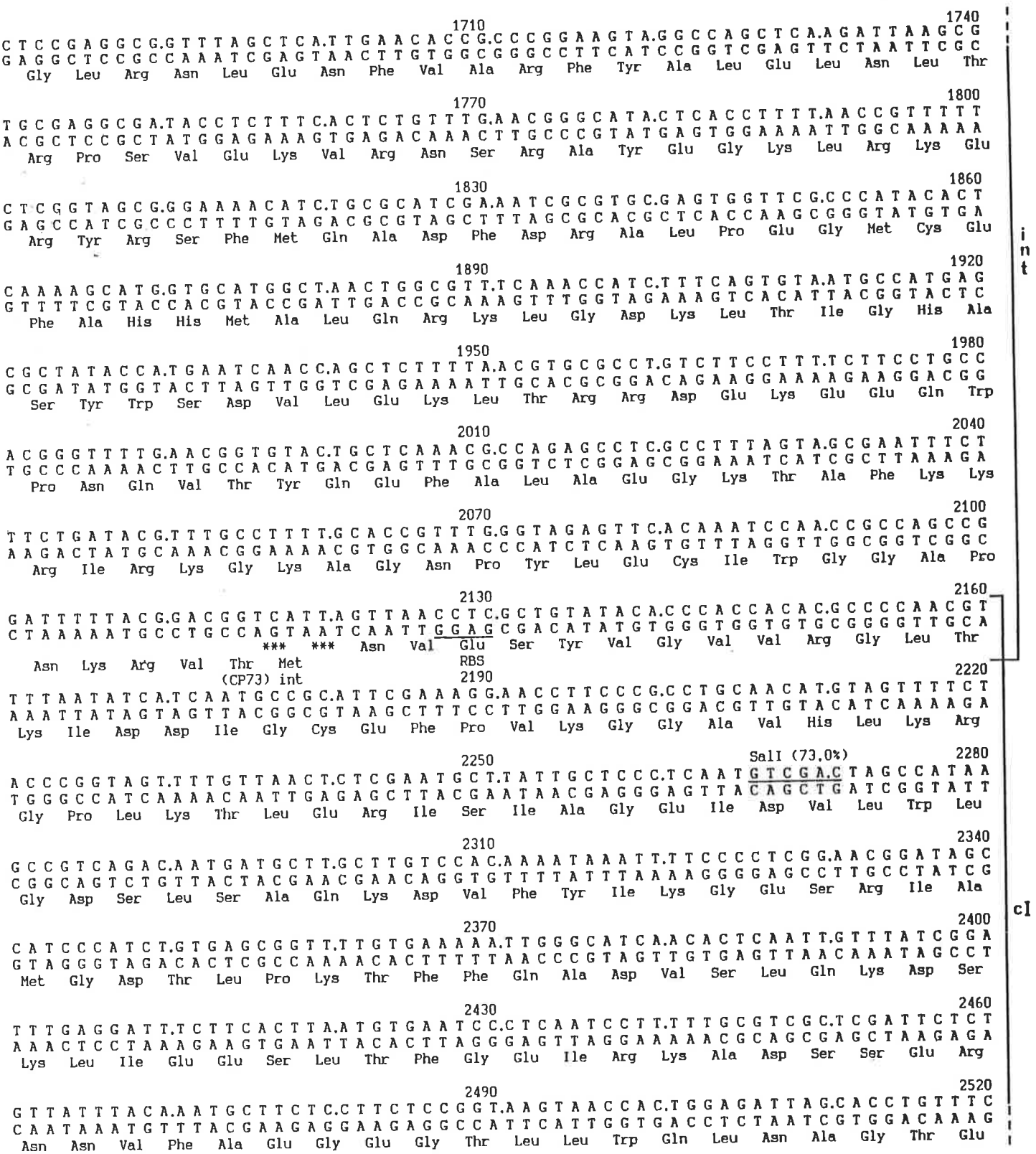
(a) Nucleotide sequence of the 3.6 kb PstI(65.5%-77.4%) restriction fragment from the DNA of 186cItsp and the amino acid sequences of proteins it most probably encodes. Transcription and translation are shown from left to right on the upper l-strand and from right to left on the lower r-strand. Numbering of the sequence is arbitrarily from the PstI(65.5%) site.

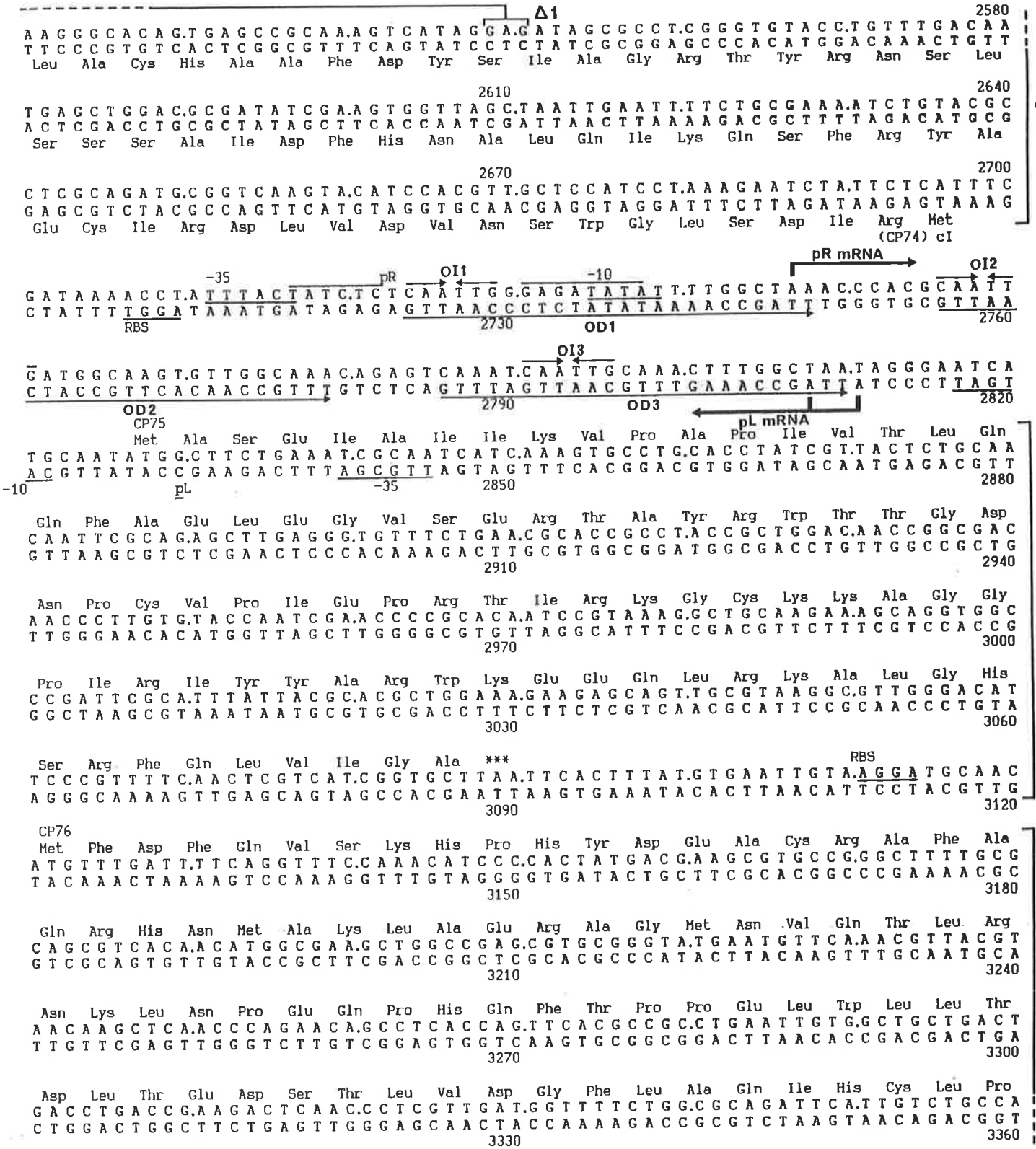
The -10 and -35 regions of promoters pB, pR and pL are underlined, as are the predicted ribosome binding sites (RBS). Direct repeats R1 and R6 are indicated by the arrows and the proposed rho-independent transcription terminators, tB and tL, are shown. Candidates for the operator sites are represented either by the direct repeats OD1, OD2 and OD3 (arrows) or the inverted repeats OI1, OI2 and OI3 (head to head arrows). Predicted startpoints for in vivo mRNA transcripts are indicated (See Section 6). Unique restriction sites on this fragment (XhoI, KpnI, SacI, SalI) are also indicated.

(b) Diagrammatic representation of the predicted coding regions. Coding regions are represented by the boxed regions and the gene assigned to each region is indicated above the box. Coding regions for genes D (CP65) and CP76 are only partly represented on the sequence.


[a]
PstI (65.5%)
D (CP65)
T Dam23
Dam14 T
D
-35
BI1
B (CP67)
B mRNA
RBS
pB
-10
BI2
B
T Bam57
T Bam17
tB
XhoI (67.6%)
tL
Δ1


R6
870
900
R5
930
R4
960
CP69
R3
990
1020
R2
1050
R1
KpnI (69.1%)
1080
1110
1140
RBS
CP69
1170
1200
1230
1260
1290
1320
1350
1380
Sac I (70.2%)
1410
1440
1470
1500
1530
1560
1590
1620
1650
1680


1710
1740
Gly Leu Arg Asn Leu Glu Asn Phe Val Ala Arg Phe Tyr Ala Leu Glu Leu Asn Leu Thr
1770
1800
Arg Pro Ser Val Glu Lys Val Arg Asn Ser Arg Ala Tyr Glu Gly Lys Leu Arg Lys Glu
1830
1860
Arg Tyr Arg Ser Phe Met Gln Ala Asp Phe Asp Arg Ala Leu Pro Glu Gly Met Cys Glu
1890
1920
Phe Ala His His Met Ala Leu Gln Arg Lys Leu Gly Asp Lys Leu Thr Ile Gly His Ala
1950
1980
Ser Tyr Trp Ser Asp Val Leu Glu Lys Leu Thr Arg Arg Asp Glu Lys Glu Glu Gln Trp
2010
2040
Pro Asn Gln Val Thr Tyr Gln Glu Phe Ala Leu Ala Glu Gly Lys Thr Ala Phe Lys Lys
2070
2100
Arg Ile Arg Lys Gly Lys Ala Gly Asn Pro Tyr Leu Glu Cys Ile Trp Gly Gly Ala Pro
2130
2160
*** *** Asn Val Glu Ser Tyr Val Gly Val Val Arg Gly Leu Thr
Asn Lys Arg Val Thr Met
RBS
(CP73) int
2190
2220
Lys Ile Asp Asp Ile Gly Cys Glu Phe Pro Val Lys Gly Gly Ala Val His Leu Lys Arg
2250
SalI (73.0%)
2280
Gly Pro Leu Lys Thr Leu Glu Arg Ile Ser Ile Ala Gly Glu Ile Asp Val Leu Trp Leu
2310
2340
Gly Asp Ser Leu Ser Ala Gln Lys Asp Val Phe Tyr Ile Lys Gly Glu Ser Arg Ile Ala
2370
2400
Met Gly Asp Thr Leu Pro Lys Thr Phe Phe Gln Ala Asp Val Ser Leu Gln Lys Asp Ser
2430
2460
Lys Leu Ile Glu Glu Ser Leu Thr Phe Gly Glu Ile Arg Lys Ala Asp Ser Ser Glu Arg
2490
2520
Asn Asn Val Phe Ala Glu Gly Glu Gly Thr Leu Leu Trp Gln Leu Asn Ala Gly Thr Glu
int
cI


Δ1
2580
2610
2640
cI
2670
2700
(CP74) cI
pR mRNA
-35
pR
OI1
-10
OI2
RBS
2730
OD1
2760
OI3
OD2
2790
OD3
2820
pL mRNA
CP75
Met Ala Ser Glu Ile Ala Ile Ile Lys Val Pro Ala Pro Ile Val Thr Leu Gln
-10
pL
-35
2850
2880
Gln Phe Ala Glu Leu Glu Gly Val Ser Glu Arg Thr Ala Tyr Arg Trp Thr Thr Gly Asp
2910
2940
CP75
Asn Pro Cys Val Pro Ile Glu Pro Arg Thr Ile Arg Lys Gly Cys Lys Lys Ala Gly Gly
2970
3000
Pro Ile Arg Ile Tyr Tyr Ala Arg Trp Lys Glu Glu Gln Leu Arg Lys Ala Leu Gly His
3030
3060
Ser Arg Phe Gln Leu Val Ile Gly Ala ***
RBS
3090
3120
CP76
CP76
Met Phe Asp Phe Gln Val Ser Lys His Pro His Tyr Asp Glu Ala Cys Arg Ala Phe Ala
3150
3180
Gln Arg His Asn Met Ala Lys Leu Ala Glu Arg Ala Gly Met Asn Val Gln Thr Leu Arg
3210
3240
Asn Lys Leu Asn Pro Glu Gln Pro His Gln Phe Thr Pro Pro Glu Leu Trp Leu Leu Thr
3270
3300
Asp Leu Thr Glu Asp Ser Thr Leu Val Asp Gly Phe Leu Ala Gln Ile His Cys Leu Pro
3330
3360

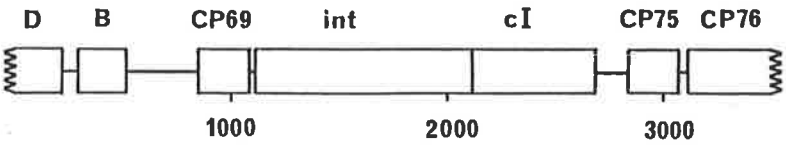
```
 Ser  Glu  Leu  Gly  Glu  Leu  Ala  Ser  Gly  Ala  Val  Ser  Asp  Glu  Arg  Leu  Thr  Thr  Ala  Arg
AGTGAACTCG.GTGAACTGGC.GAGCGGGTGCG.GTATCTGATG.AGCGTCTGAC.CACTGCCCGT
TCACTTGAGCCACTTGACCGCTCGCCCACGCCATAGACTACTCGCAGACTGGTGACGGGCA
                                  3450                          3480

 Lys  His  Asn  Met  Ile  Glu  Ser  Val  Asn  Ser  Gly  Ile  Arg  Met  Leu  Ser  Leu  Ser  Ala  Leu
AAGCACAACA.TGATTGAAAG.CGTTAACTCC.GGCATTCGCA.TGTTGTCATT.GTCGGCTCTG
TTCGTGTTGTACTAACTTTCGCAATTGAGGCCGTAAGCGTACAACAGTAACAGCCGAGAC
                                  3510                          3540

 Ala  Leu  His  Ala  Arg  Leu  Gln
GCGCTGCAT G.CACGTCTGCA.G
CGCGACGTA CGTGCAGACGTC
                     3561
                PstI (77.4%)
```

C P 7 6

[b]



SECTION 5

IDENTIFICATION OF POTENTIAL GENES

5. IDENTIFICATION OF POTENTIAL GENES

5.1 INTRODUCTION

A knowledge of the organisation and function of the genes on the DNA sequence is an essential prerequisite to understanding gene control mechanisms. Genetic mapping studies (Hocking and Egan, 1982a; Finnegan and Egan, 1979) placed genes cI, int, B and part of the D gene on the PstI(65.5%-77.4%) fragment and the aim of this section is, where possible, to provide evidence that identifies the genes on this DNA fragment. As a direct consequence of the availability of the completed DNA sequence, several studies were initiated by other workers and unpublished data from these studies which provides evidence to identify genes on the DNA sequence will be used, with due reference given to the workers involved.

Positive identification of a gene can be made by locating the single base-pair mutation that generates an amber translation termination codon in the reading frame of the gene of interest, thereby confirming both the open reading frame and the direction of transcription for that gene. Amber mutants genes D, B and cI have been reported (Hocking and Egan, 1982a), however the DNA sequences of amber mutants in genes B and cI were determined by other workers (Section 5.2.2, 5.2.3). To complete the analysis of amber mutants known to be located on the PstI(65.5%-77.4%) fragment, two alleles of the D gene (Dam23, Dam13) which map to the immediate right of the PstI site at 65.5% (Finnegan and Egan, 1979), were sequenced.

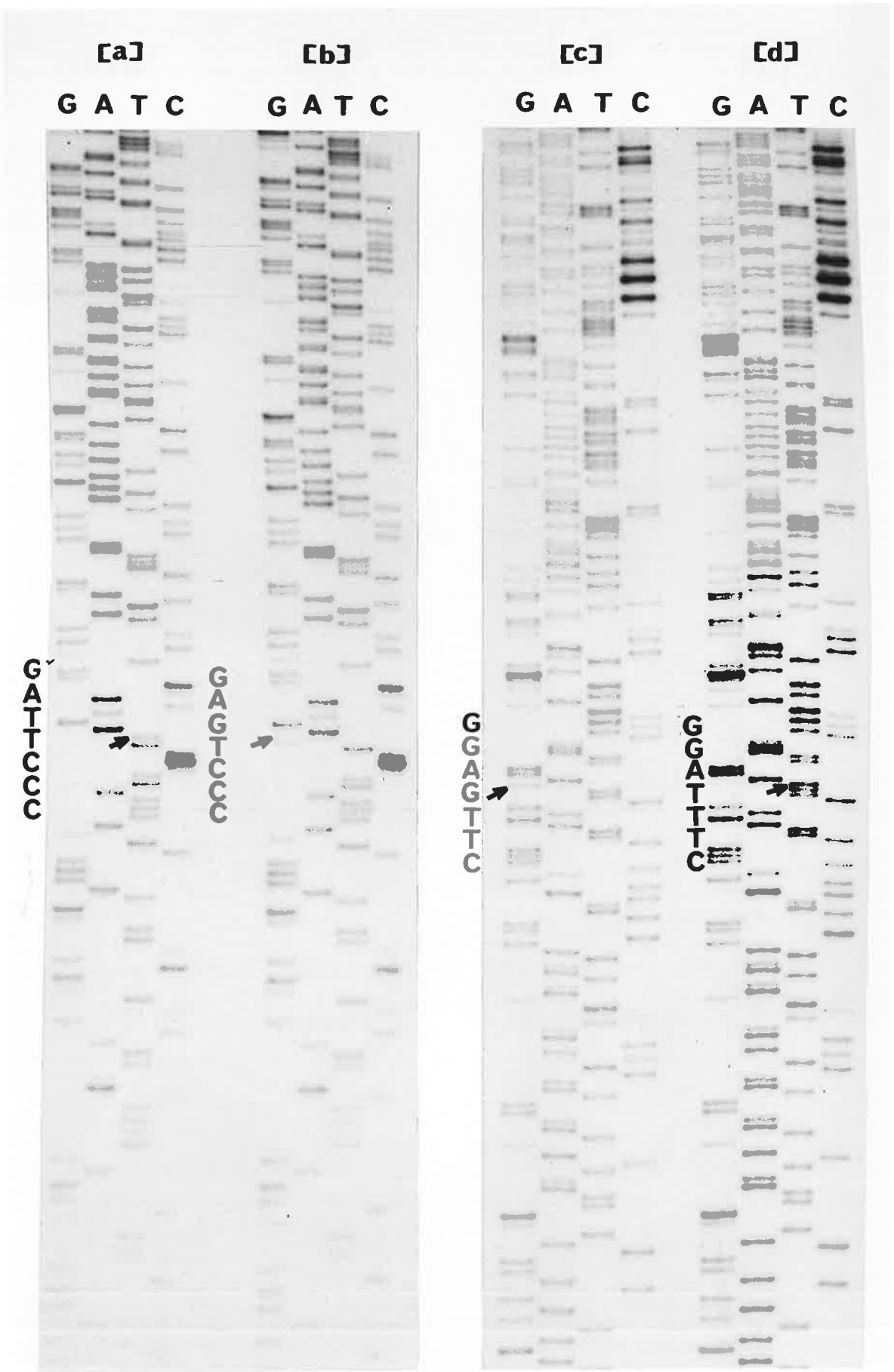
5.2 RESULTS AND DISCUSSION

5.2.1 D Gene

DNA from each amber mutant was digested with PstI and XhoI, and the mixture of DNA fragments was force-cloned into M13mp8 vector DNA which had been restricted with PstI and SalI, to ensure that the sequence of each clone would be determined from the PstI site.

Fig. 5.1. DNA sequence of mutants Dam23 and Dam14

Dam23 and Dam14 DNA (10 ug) was doubly-digested with PstI and XhoI, and the resulting fragments cloned into M13mp8 doubly-digested with PstI and SalI. M13 clones containing the 0.6 kb PstI-XhoI(65.5%-67.6%) fragment were isolated (Section 5.2.1) and sequenced. The DNA sequence of mutants (a) Dam23 and (d) Dam14 from the PstI(65.5%) site is presented, and alongside each is the DNA sequence of 186cItsp DNA ((b) and (c)) for comparison. All sequences are those of the l-strand. Single base-pair changes that generate amber termination codons in Dam23 and Dam14 were located as indicated by the arrows in (a) and (d) respectively. Electrophoresis of samples in (a) and (b) was for 90 minutes whilst the samples in (c) and (d) were electrophoresed for 2.5 hours. Electrophoresis and autoradiography conditions were described in Sections 2.18.4(b) and 2.18.5.


[a]
G A T C
GATTCCC
[b]
G A T C
GAGTCCC
[c]
G A T C
GGAGTTC
[d]
G A T C
GGATTTC

Two different DNA fragments could ligate into this vector (2.9 kb and 0.6 kb PstI/XhoI restriction fragments) and single-stranded clones containing the smaller 0.6 kb insert were selected on the basis of size and sequenced (Section 2.19.8, 2.20).

Nucleotide sequences from the PstI(65.5%) site toward the XhoI(67.6%) site of Dam23, Dam14 and 186am^{+}DNA are compared in Figure 5.1. Mutants Dam23 and Dam14 were both found to be GAG to TAG transversions (Fig. 5.1) at sequence positions 64 and 178 respectively. An amber termination codon in the reading frame of CP65 was generated in each case, identifying CP65 as the D gene and confirming its transcription rightward relative to the genetic map. Two alleles of the D gene (Dam26 and Dam48) map to the left of the PstI(65.5%) site (Finnegan and Egan, 1979), indicating that gene D is only partially represented on the DNA sequence of the PstI(65.5%–77.4%) fragment.

5.2.2 B Gene

The DNA sequence of mutants Bam57 and Bam17 was determined by Pritchard (1984) and compared to the DNA sequence of the PstI (65.5%–77.4%) fragment. The Bam57 mutation was found to be a CAG to TAG transition at sequence position 379 and the Bam17 mutation was a GAG to TAG transversion at sequence position 430. In each case an amber translation termination codon was generated in the reading frame of CP67, identifying CP67 as the B gene and confirming its transcription rightward relative to the genetic map. Furthermore, the position of the amber mutations found in the D and B genes agreed with the map order Dam23–Dam14–Bam57–Bam17, published from the frequency of marker rescue data (Hocking and Egan, 1982a).

5.2.3. cI Gene

Insertion mutant ins1 gives the cI$^-$ phenotype and is located at position 73.5±0.7% (Younghusband et al., 1975). CP74 was found to span this region (74.5%-72.6%) and is predicted to be the cI gene.

Positive identification of this gene was provided by the DNA sequence of the cIam461 mutant (I. Lamont, personal communication). This mutation was found to be a TGG to TAG transition at sequence position 2675, generating an amber translation termination codon in CP74. The mutation identified CP74 as the cI gene and confirmed that its transcription was leftward relative to the genetic map.

Two possible translational initiation codons were identified for the reading frame of the cI (CP74) gene (Section 4.2.2). The location of the cIam461 mutation was between these two initiation codons, which precluded the possibility that the GTG initiation codon at sequence positions 2667 to 2665 was the translational initiation codon for the cI gene. The most likely initiation codon (sequence co-ordinates 2697-2695) and ribosome binding site for the cI gene are presented in Table 4.1.

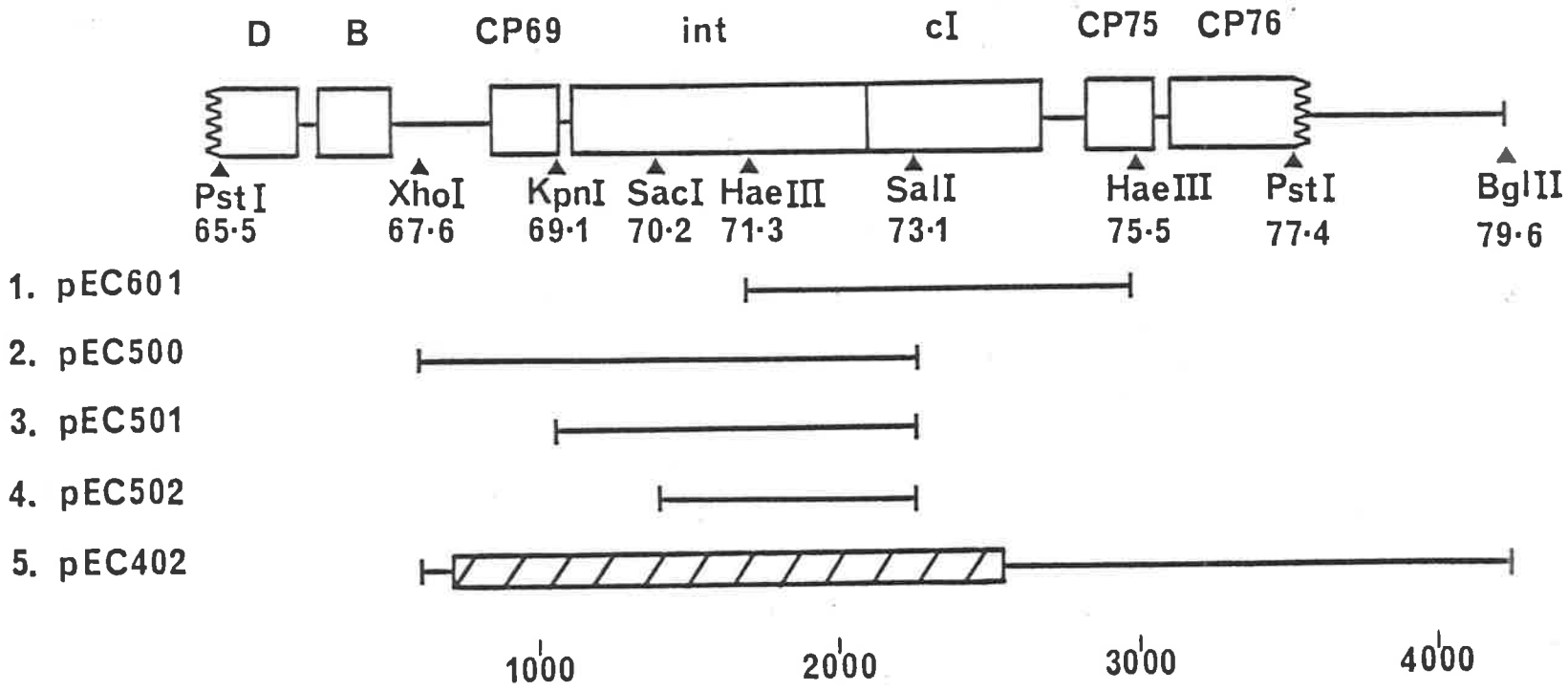
A plasmid clone (pEC601) of the 1276 bp HaeIII(71.2%-75.5%) fragment, spanning the cI gene (Fig. 5.2 and sequence co-ordinates 1723-2999, was used in the maxicell system described by Sancar et al. (1979). Translation products directed by this plasmid clone were analysed by gel electrophoresis and a protein of 22.7 Kd was identified as a potential candidate for the cI protein (A. Puspurs, unpublished data). The size of this protein was in good agreement with the 21.2 Kd product predicted by the amino acid translation of the cI gene from the DNA sequence (Table 8.1).

Fig. 5.2. Plasmid clones used to identify potential phage 186 protein products

The DNA fragment from the PstI(65.5%) site to the BglII(79.6%) site is shown diagrammatically, with the proposed genes on the PstI(65.5%-77.4%) fragment indicated. Restriction sites used to generate plasmid clones and their percentage positions on the 186 genome are shown.

Plasmid clones were constructed from vectors pPLc236, pKC7 (Remaut et al., 1981; Nagaraja and Rogers, 1979) or a derivative of vector pKO1 (McKenney et al., 1981) named pKO1-T (gift from K.M. McKenney). DNA restriction fragments cloned into these vectors are indicated by the bars.

1. pEC601, a pKO1-T clone of the HaeIII(71.3%-75.5%) DNA fragment (constructed by I.B. Dodd)
2. pEC500, a pPLc236 clone of the XhoI-SalI(67.6%-73.1%) DNA fragment (constructed by A. Puspurs).
3. pEC501, a pPLc236 clone of the KpnI-SalI(69.1%-73.1%) DNA fragment (constructed by A. Puspurs).
4. pEC502, a pPLc236 clone of the SacI-SalI(70.2%-73.1%) DNA fragment (constructed by A. Puspurs).
5. pEC402, a pKC7 clone of the XhoI-BglII(67.6%-79.6%) DNA fragment which contains the Δ1 DNA deletion (indicated by the shaded box). The Δ1 deletion extends from 67.9% to 74.0% (Section 4.2.7).


D
B
CP69
int
cI
CP75
CP76
PstI
65·5
XhoI
67·6
KpnI
69·1
SacI
70·2
HaeIII
71·3
SalI
73·1
HaeIII
75·5
PstI
77·4
BglII
79·6
1. pEC601
2. pEC500
3. pEC501
4. pEC502
5. pEC402
1000
2000
3000
4000

5.2.4 Int Gene

Insertion mutant ins3 gives the Int$^-$ phenotype and is located at position 70.3%±0.7% (Younghusband et al., 1975). CP73 was found to span this region (72.6%-69.2%) and is predicted to be the int gene. Later discussion (Section 8.3) will detail the considerable homology between CP73 and the known P2 int gene which provides further evidence that CP73 is the int gene of phage 186.

Deletion mutant Δ1 displays the Int cI mutant phenotype (Dharmarjah, 1975) and the DNA deletion extends from 69.7% to 74.0% of the 186 genome (Section 4.7), which approximates the int-cI region. The assignment of CP73 and CP74 as the int and cI genes respectively, was consistent with the properties of the Δ1 mutant.

The size of the int gene product predicted by translation of the DNA sequence was 38.9 Kd (Table 8.1). In the maxicell system (Sancar et al., 1979), a plasmid clone (pEC500) of the XhoI(67.6%) to SalI(73.1%) region (Fig. 5.2) directed the synthesis of two proteins which were sized from the gel at 38.9 Kd and 8.4 Kd. (A. Puspurs, personal communication). Only the 38.9 Kd protein was produced from a plasmid clone (pEC501) of the KpnI(69.1%) to SalI(73.1%) region (Fig. 5.2) and furthermore, the 38.9 Kd protein disappeared when the plasmid clone (pEC502) of the SacI(70.2%) to SalI(73.1%) region (Fig. 5.2) was used (A. Puspurs, personal communication). These results are consistent with the prediction that the int (CP73) gene was located in the DNA interval between the KpnI(69.1%) and SalI(73.6%) sites and encoded a protein product of 38.9 Kd (Table 8.1).

5.2.5 CP69, CP75 and CP76

Genetic studies to date, have failed to detect mutants in CP69, CP75 or CP76 (only partly represented) which would positively identify them as genes of phage 186. Some evidence for the existence of these genes comes from the identification of proteins in the maxicell system (Sancar et al., 1979) which correspond in size to that predicted by computer-translation of these genes using the MWCALC program (Staden, 1980).

A plasmid clone (pEC402) of the XhoI(67.6%) to BglII(79.6%) region containing the Δ1 deletion (Fig. 5.2), directed the synthesis of several phage-specific proteins in the maxicell system (Sancar et al., 1979), one of which was sized at 9.8 Kd (A. Puspurs, personal communication. This protein was considered a potential candidate for the 9.7 Kd CP75 gene-product (Table 8.1). A potential candidate for CP76 has also been identified (See Section 8.5).

The gene-product of CP69 was predicted to be a protein of 8.4 Kd by amino acid translation of the DNA sequence (Table 8.1). Two phage-specific proteins (8.4 Kd and 38.9 Kd) were described in Section 5.2.4, which were synthesised from the plasmid clone (pEC500) of the XhoI(67.6%) to SalI(73.1%) region (Fig. 5.2). The 8.4 Kd phage-specific protein was not synthesised from the plasmid clone (pEC501) of the KpnI(69.1% to SalI(73.1%) region (Fig. 5.2) (A. Puspurs, personal communication). CP69 (69.2%-68.3%) spans the KpnI(69.1%) site, suggesting that the 8.4 Kd phage-specific protein product synthesised by the plasmid clone pEC500 is a likely candidate for the CP69 gene product.

SECTION 6

CONFIRMATION OF PROMOTERS ON THE PstI(65.5%-77.4%) FRAGMENT

BY ANALYSIS OF RNA TRANSCRIPTS IN VIVO

6. CONFIRMATION OF PROMOTERS ON THE PstI(65.5%-77.4%) FRAGMENT BY ANALYSIS OF RNA TRANSCRIPTS IN VIVO

6.1 INTRODUCTION

Computer analysis of the DNA sequence was used to predict the location of potential transcription promoters and terminators. The limitation of this analysis is that positively controlled promoters could well escape detection, since they would not be expected to show homology to the E.coli consensus promoter. Three potential promoters were located in the DNA sequence of the PstI(65.5%-77.4%) fragment which were named pB, pL and pR (Fig. 4.5 and Table 4.2).

Promoter pB was found in the sequence preceding the B gene and would initiate rightward transcription of the B gene. Transcription from pB would terminate at the proposed rho-independent terminator tB and a transcript of approximately 290 b is predicted. A transcript of this size was identified in vitro and positively identified as the B gene mRNA transcript by determining the RNA sequence of the 290 b transcript from mutants Bam17 and Bam57 (Pritchard 1984; Pritchard and Egan, 1985). Although several inverted repeats were found in the DNA sequence, between the tail gene D and the B gene, which may represent rho-dependent transcription terminators, no rho-independent terminators were located in this sequence. The possibility therefore existed that the B gene could be expressed from a larger transcript late in infection.

Promoter pL was located in the sequence preceding the cI gene and would initiate leftward transcription of cI, int and CP69 and terminate at the rho-independent terminator tL. Assuming that initiation of this mRNA occurs 4 to 7 b after the -10 hexamer of pL (See Rosenberg and Court, 1979), a leader sequence of 120 b is expected before the cI gene, and a total length of approximately 2.2 kb is predicted for the pL transcript. Transcription of the

att-int-cI region was shown to be dependent on protein synthesis (Finnegan and Egan, 1981) and it was therefore not unexpected that the predicted pL transcript was not identified as one of the four major in vitro transcripts.

Promoter pR was predicted to initiate rightward transcription of CP75 and CP76, and extend past the PstI site at 77.4%. The 5'-end of the 1.45 kb in vitro RNA transcript was predicted to correspond to the A residue at sequence position 2747, consistent with a transcript initiating from the pR promoter (the -10 region of the pR promoter is located at sequence co-ordinates 2735-2740) and the 1450 b transcript therefore extends some 600 b past the PstI site at 77.4% (sequence position 3561; Pritchard, 1984; Pritchard and Egan, 1985). Finnegan and Egan (1981) provided evidence that the region from 74.9% to 77.4% included the early lytic transcript and that transcription of this region was under cI repressor control. The 1.45 kb in vitro transcript initiating from the pR promoter (named the pR transcript) was therefore predicted to be the early lytic transcript of phage 186 that is under cI repressor control.

Promoters pL and pR are predicted to initiate transcription of the lysogenic and lytic transcripts respectively, but the two promoters show a different arrangement to the analogous promoters in phage λ (λpRM and pR). Transcription from the two 186 promoters initially converges and then diverges (named convergent overlapping). Since the two promoters are predicted to be crucial in the decision to follow the lytic or lysogenic pathways, it was essential to confirm the activity of both promoters in vivo.

The primary aim of this section was to confirm the in vivo activity of promoters pB, pR and pL by identifying the 5'-end of the associated RNA transcripts. The obvious assumption in these studies was that the 5'-ends of the transcripts detected were true

RNA transcription startpoints rather than the products of RNA processing or degradation. Northern analysis was used to identify and size RNA transcripts (Section 1.7.2).

6.2 RESULTS AND DISCUSSION

6.2.1 Transcription from the pB Promoter

6.2.1(a) Identification of the B Gene Transcript

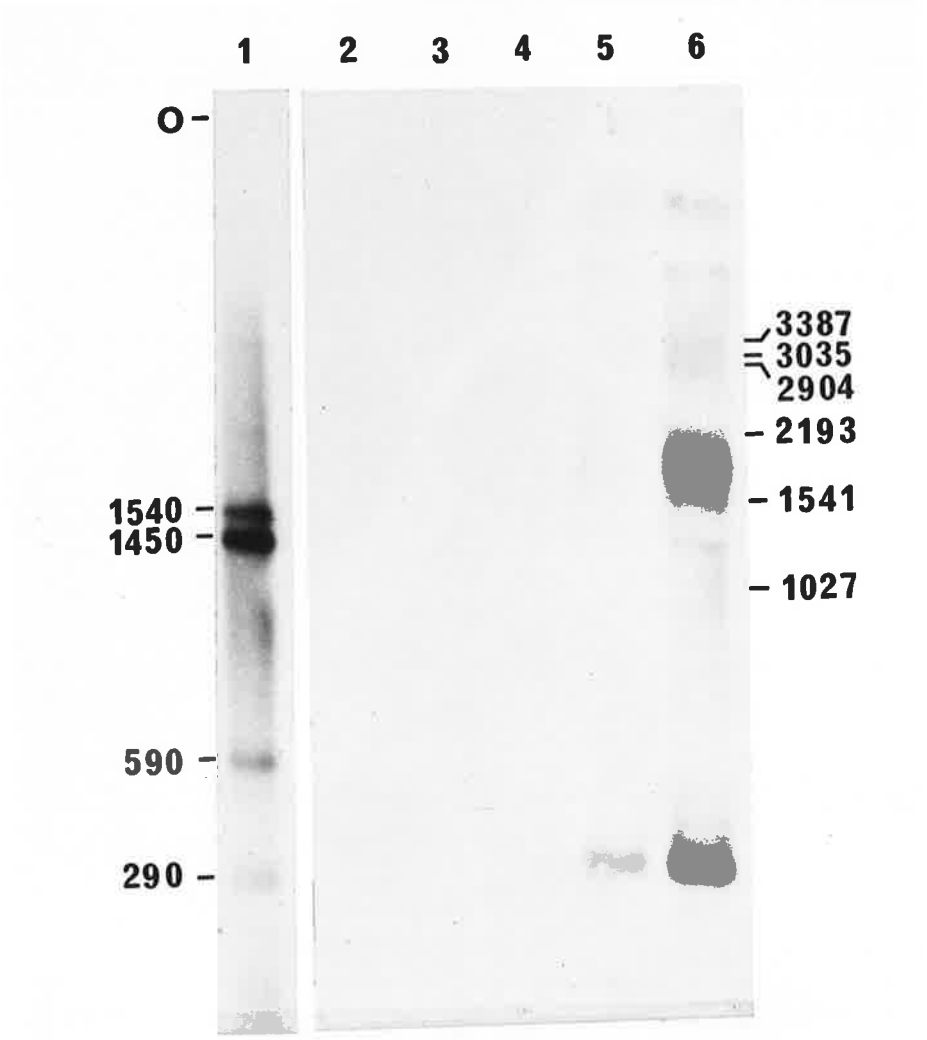
An M13 single-stranded clone of the PstI-XhoI(65.5%-67.6%) fragment, with the same polarity as the B mRNA, was used to prepare a B-specific probe. The 17-mer universal primer (Section 2.6) was annealed to this clone, extended across the insert, and the 100 bp radioactive HhaI fragment (sequence co-ordinates 334-434) was isolated after restriction digestion and fractionation of the resulting fragments on a 5% polyacrylamide gel, as described in Section 2.21. Detection of mRNA with this radioactive probe would automatically indicate rightward transcription.

B gene mRNA synthesis during vegetative development was investigated by hybridising the B-specific probe to total cellular RNA, isolated at different times after heat induction of a 186cItsp prophage, using the Northern analysis procedure described in Section 2.22.4. The B-specific probe detected an *in vivo* transcript which comigrated with the 290 b *in vitro* B transcript (Fig. 6.1) and the amount of the 290 b *in vivo* transcript increased dramatically in the samples taken 25 and 35 minutes after heat induction (Fig. 6.1, tracks 5 and 6). The 25 minute time period corresponds to the onset of DNA replication and the transition to late gene transcription (Hocking and Egan, 1982c; Finnegan and Egan, 1981). Longer exposure of the autoradiograph revealed that the 290 b *in vivo* transcript could be detected in the sample taken 5 minutes after prophage induction (data not shown).

Larger transcripts were also detected with the B specific

Fig. 6.1. B gene transcripts after prophage induction

E251(186cItsp) was heat induced (0 min) and, at increasing times thereafter, 10 ml samples were collected and RNA recovered (Section 2.22.1). 100 ug of RNA was denatured with glyoxal, fractionated on a 1.8% horizontal agarose gel, transferred bi-directionally to nitrocellulose (Section 2.22.4) and probed with the B specific probe (3-5 x 10^6 cpm) (See Fig. 6.2(a)). Track 1 is a marker track of 186 *in vitro* transcripts (1540 b, 1450 b, 590 b, 290 b). Non-radioactive *E. coli* 23S RNA (2904 b), 16S RNA (1541 b) and CMV markers (3387 b, 3035 b, 2193 b, 1027 b) are also indicated in the figure. Track 2: "Induction" of the non-lysogen E251, 0-35 min sample. Tracks 3-6: Induction of E251(186cItsp) with samples 0-5 min (Track 3) 0-15 min (Track 4), 0-25 min (Track 5) and 0-35 min (Track 6). The autoradiograph was exposed for 2 days at -80°C with an intensifying screen.


1
2
3
4
5
6
O
1540
1450
590
290
3387
3035
2904
2193
1541
1027

probe; most noticeable was a transcript of about 2 kb which was detected in the 25 and 35 minute time samples (Fig. 6.1, tracks 5 and 6).

Preliminary Northern analysis suggested that the 2 kb transcript could also be detected with a nick translated BamHI-HindIII (58.7%-61.3%) probe from the late region (data not shown). The BamHI-HindIII fragment is about 1.2 kb from the PstI site at 65.5%. The appearance of the 2 kb transcript in the late period after induction and its detection using a probe from the late region preceding the B gene, indicates that it is a late gene transcript.

6.2.1(b) Sequence of the 5'-end of the B Gene Transcript

To confirm that the 290 b transcript detected by the B-specific probe did initiate from the pB promoter, the B-specific probe was used as a radioactive primer to determine the 5'-terminal sequence of 290 b transcript, as described in Section 2.22.6. In this method a radioactive primer is annealed to total cellular RNA in liquid and the hybrid(s) (also called primed-RNA species) are fractionated on a non-denaturing gel. The primed-RNA species are detected by autoradiography, excised from the gel and eluted from the gel slice. The dideoxynucleotide chain termination technique using AMV reverse transcriptase was used to determine the sequence from the primer. Fractionation of the primed-RNA species was carried out to prevent overloading of the sequencing gels with RNA which results in poor quality sequences (R.H. Williams, personal communication).

Northern analysis revealed that the B-specific probe hybridised to transcripts of 290 b and 2 kb, which would potentially create problems in identifying the 5'-end of the 290 b B gene transcript by sequencing, since priming would take place on both transcripts. Conditions were therefore sought where transcription of only the 290 b transcript occurred. Transcriptional activity of the B gene

region was shown to be markedly enhanced after heat induction of a Bam17cItsp lysogen but the transition to late gene transcription did not occur (Finnegan and Egan, 1981; Section 1.5.2). The 2 kb transcript detected by Northern analysis with the B-specific probe appeared late in infection (Fig. 6.1, tracks 5 and 6) and most probably represents a late transcript (see also Section 6.2.1(a)). RNA isolated after induction of a Bam17cItsp lysogen should therefore be a source of the 290 b transcript alone.

Total cellular RNA isolated 25 minutes after heat induction of a Bam17cItsp lysogen was added to the radioactive B-specific primer (100 bp HhaI fragment, see Fig. 6.2(a)). The solution was heat denatured at 90°C for 5 minutes and the primer annealed for 4.5 hours at 45°C. The resulting primed-RNA species were fractionated on a 5% non-denaturing polyacrylamide gel (Fig. 6.2(b)). Two bands were detected (Fig. 6.2, track 2), one of which was the size of the unannealed primer in the control track (Fig. 6.2(b), track 1); the other was predicted to be the primed-RNA species. This was confirmed by excising the band containing the primed-RNA species from the gel and sequencing to the 5'-end of the RNA from the B-specific primer (Section 2.22.6(b)).

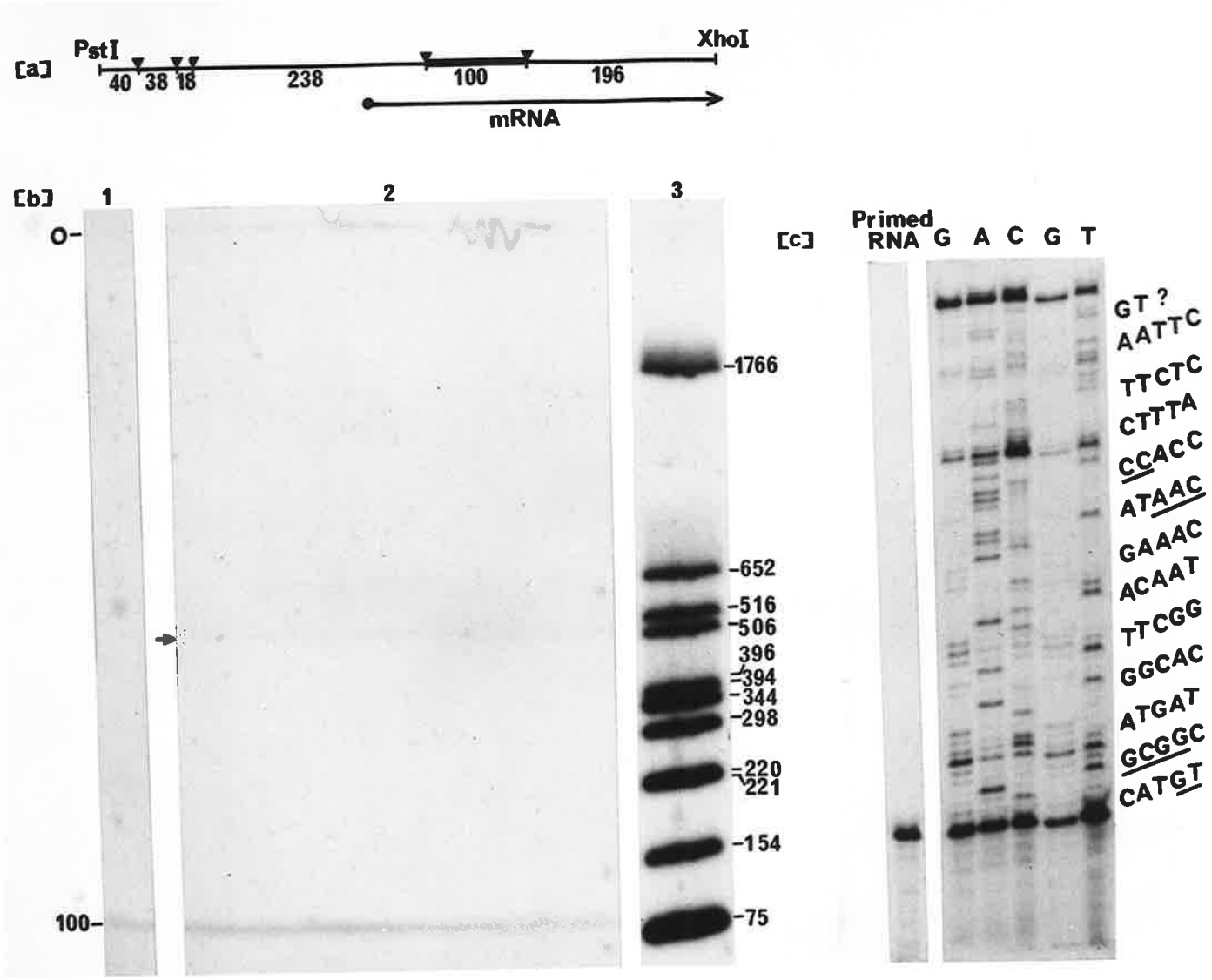
Figure 6.2(c) is the DNA sequence gained by extension of the B-specific primer. This sequence corresponds to the 5'-terminal sequence of the 290 b *in vitro* B gene transcript determined by direct RNA sequencing (Pritchard, 1984; Pritchard and Egan, 1985). The sequence ladder is cross-banded in several places and this problem is particularly noticeable when the AMV reverse transcriptase encounters the 5'-end of the template, preventing the determination of exact 5'-terminal residue. A similar problem was reported by Akusjarvi and Pettersson (1979) using essentially the same sequencing procedure. Presumably the B gene transcript *in vivo* starts at the

Fig. 6.2. Sequence of the 5'-end of the B gene transcript

(a) Restriction map of the PstI-XhoI(65.5%-67.6%) DNA fragment showing the sizes of the fragments generated by HhaI (▼) digestion. The 100 b HhaI primer is shown and its position relative to the predicted startpoint of the B mRNA is also indicated (Section 6.1).

(b) E251(186cItsBam17) was heat induced at 39°C for 25 min and RNA isolated from a 10 ml aliquot. The radioactive B specific primer ($2-3 \times 10^5$ cpm) (Section 6.2.1(a)), prepared as described in Section 2.21, was added to 100 ug RNA, precipitated with ethanol and resuspended in 20 ul of 70% de-ionised formamide, 66 mM KH_2PO_4 (pH 7.0). The mixture was heated at 90°C for 5 min, followed by 4.5 hours incubation at 45°C and then fractionated on a 5% polyacrylamide gel in TBE. Track 1 is a control track containing 10 ug RNA and radioactive primer which were heated to 90°C and electrophoresed immediately to prevent annealing. The 100 b primer is indicated in Track 1. Track 2 contains 100 ug RNA annealed to the radioactive primer for 4.5 hours at 45°C. The primed-RNA species is arrowed and unannealed primer is also present. Track 3 is a marker track containing pBR325 DNA digested with HinfI and end-labelled (Section 2.19.4). The autoradiograph was exposed overnight at room temperature.

(c) The single primed-RNA species, detected by autoradiography, was eluted overnight (Section 2.18.4(d)). Nucleic acid was phenol extracted and precipitated with ethanol. The sequence was determined using AMV reverse transcriptase as described in Section 2.22.6(b). Sequences which are compressed are underlined and the primed-RNA species is also shown. The actual start point of the mRNA cannot be conclusively determined due to the compressed band at the top of the sequence, but is presumed to start with the A residue at sequence position 270 in Figure 4.5, which corresponds to the T? residue. Electrophoresis conditions were described in Section 2.18.4(b) and autoradiography was for one week at -80°C with an intensifying screen.


[a]
PstI
XhoI
40 38 18
238
100
196
mRNA
[b]
1
2
3
O-
100-
1766
652
516
506
396
394
344
298
220
221
154
75
[c]
Primed RNA
G A C G T
GT ?
AATTC
TTCTC
CTTTA
CCACC
ATAAC
GAAAC
ACAAT
TTCGG
GGCAC
ATGAT
GCGGC
CATGT

same A residue (sequence position 270 which corresponds to the T residue in Fig. 6.2(c)) predicted to be the startpoint of the B gene transcript *in vitro* (Pritchard, 1984; Pritchard and Egan, 1985). Associated with the 5'-end of the 290 b B gene transcript is the -10 region of the pB promoter (sequence co-ordinates 258-263), which suggests that this transcript initiates from the pB promoter both *in vivo* and *in vitro*.

6.2.2. Transcription from the pR Promoter

6.2.2(a) Identification of the pR Transcript

DNA sequence studies predicted the promoter for early lytic gene transcription (pR) was located at sequence position 2712 to 2740 (see Section 4.2.3). A 1.45 kb transcript was one of the four major *in vitro* transcripts, the 5'-end of this transcript was determined by sequencing from a radioactive primer (using the same method described in Section 2.22.6) and was found to be consistent with a transcript initiating from the pR promoter (Pritchard, 1984; Pritchard and Egan, 1985).

Essentially the same radioactive primer was used as a probe by H. Richardson of this laboratory, to detect *in vivo* mRNA transcripts synthesised after induction of a 186cItsp prophage by Northern analysis. A 1.45 kb band was detected, which comigrated with the *in vitro* pR transcript in the marker track but several other bands were also detected which were sized at 1.1 kb, 0.87 kb, 0.77 kb and 0.61 kb (H. Richardson, B. Kalionis and J.B. Egan, manuscript in preparation). Although the detection of a 1.45 kb *in vivo* transcript provides evidence that the pR promoter is active *in vivo*, more conclusive evidence would be to identify the 5'-end of the transcript(s) initiating from this region.

6.2.2(b) Sequence of the 5'-end of the pR Transcript

To ascertain whether RNA transcription initiated from the pR promoter *in vivo*, essentially the same primer used to determine the sequence of the 5'-end of the 1.45 kb *in vitro* pR transcript was annealed to *in vivo* mRNA and the sequence from the primer to the 5'-end of the *in vivo* transcript(s) was determined.

The radioactive primer was prepared by synthesising a radioactive copy of an M13 single-stranded HpaII clone (sequence co-ordinates 2489 to 2935) from the universal 17-mer primer (Section 2.21). After digestion with HinfI and BamHI and fractionation of the resulting fragments on a 5% polyacrylamide gel, the 124 bp BamHI/HinfI radioactive fragment (117 bp of 186 DNA from HinfI to HpaII, sequence co-ordinates 2818-2935 and 7bp to the BamHI cloning site of M13mp7, Fig. 6.3(a)) was isolated.

RNA isolated 25 minutes after heat induction of a Bam17cItsp lysogen was used as a source of 186 *in vivo* mRNA since transcription of the early region was shown to be markedly enhanced when the B protein was defective (Finnegan and Egan, 1981; see also Section 1.5.2). Annealing of the radioactive primer to the RNA and the subsequent fractionation, isolation and sequencing of the primed-RNA species was as described in Section 6.2.1(b)).

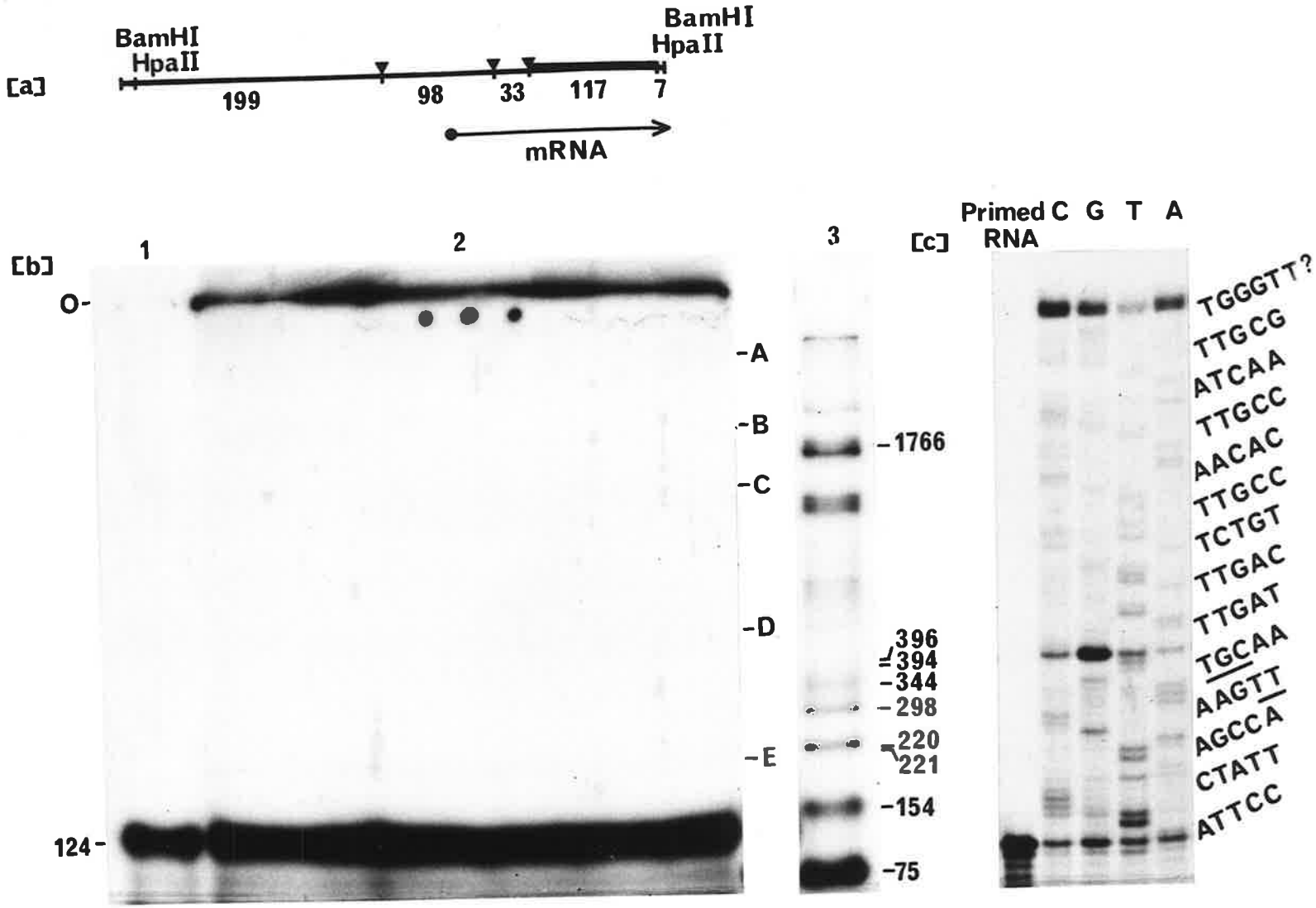
Five major primed-RNA species (A-E) were detected after annealing of the radioactive primer to Bam17cItsp RNA (Fig. 6.3(b)), and the sequence generated by extending the radioactive primer on primed-RNA species A is presented in (Fig. 6.3(c)). Sequences generated by extending the radioactive primer on the other four primed-RNA species (B-E) were identical to that of primed-RNA species A (data not shown). The sequence in Figure 6.3 (c) is identical to that generated from the EcoRI/HinfI radioactive primer on the pR transcript *in vitro* (Pritchard 1984; Prichard and Egan, 1985).

Fig. 6.3. Sequence of the 5'-end of the pR transcript

(a) Restriction map of the HpaII fragment (sequence positions 2489-2935) showing the sizes of the fragments generated by HinfI (▼) digestion. The BamHI sites are located in the M13mp7 polylinker site (See Section 6.2.2(b)). The 124 b BamHI/HinfI primer is shown and contains 117 b of 186 DNA (thick line) and 7 b of M13mp7 DNA to the BamHI site. Also indicated is the position of the primer relative to the predicted startpoint of the pR mRNA (Section 6.1).

(b) Isolation of RNA after induction of E251(186cItspBam17), annealing of the primer and fractionation of the primed-RNA species were as described in the legend to Figure 6.2. Track 1 is a control track containing 10 ug RNA and radioactive primer, which were heated to 90°C and electrophoresed immediately to prevent annealing. The 124 b primer is indicated in Track 1. Track 2 contains 100 ug RNA annealed to the radioactive primer for 4.5 hours at 45°C. The primed-RNA species A-E are indicated and unannealed primer is also present. Track 3 is a marker track containing pBR325 DNA digested with HinfI and end-labelled (Section 2.19.4). The digest did not go to completion and prevented the identification of all the HinfI fragments.

(c) Each of the primed species A-E, detected by autoradiography, was eluted overnight (Section 2.18.4(d)). Nucleic acid was phenol extracted and precipitated with ethanol. The sequence was determined using AMV reverse transcriptase as described in Section 2.22.6(b). The sequence generated from primed-species A is presented and sequences generated from primed-species B-E were identical (data now shown). The actual startpoint of the mRNA cannot be conclusively determined due to the compressed band at the top of the sequence, but is presumed to start with A residue at sequence position 2747 in Figure 4.5, which corresponds to the T? residue. Electrophoresis conditions were described in Section 2.18.4(b) and autoradiography was for two weeks at -80°C with an intensifying screen.


[a]
BamHI
HpaII
199
98
33
117
7
BamHI
HpaII
mRNA
[b]
1
2
3
O-
124-
-A
-B
-C
-D
-E
-1766
396
394
-344
-298
220
221
-154
-75
[c]
Primed RNA
C G T A
TGGGTT?
TTGCG
ATCAA
TTGCC
AACAC
TTGCC
TCTGT
TTGAC
TTGAT
TGCAA
AAGTT
AGCCA
CTATT
ATTCC

The two primers differ only in the restriction enzyme (EcoRI or BamHI) used to excise the primer from the M13mp7 vector DNA. As described previously in Section 6.2.1(b), sequences generated in this manner tend to be cross-banded, particularly when the AMV reverse transcriptase encounters the end of the template. Although the DNA sequence appears to terminate with TTT residues (corresponding to AAA residues at sequence positions 2747 to 2749) there is some doubt about the final T residue. The 5'-end of the *in vivo* RNA transcript was assumed to be identical to that of the *in vitro* pR transcript (Pritchard, 1984; Pritchard and Egan, 1985), and corresponds to the A residue at sequence position 2747. Associated with this 5'-end is the pR promoter (the -10 region of the pR promoter is at sequence co-ordinates 2735 - 2740) which suggests that the pR promoter is an active promoter *in vivo* as well as *in vitro*.

The occurrence of five primed-RNA species requires some comment on their possible nature. The primed-RNA species all had a common 5'-end when sequenced. These primed-RNA species may result from annealing of the primer not only to the 1.45 kb *in vivo* transcript detected by Northern analysis (which is presumed to be the equivalent of the 1.45 kb pR transcript), but also to the other *in vivo* transcripts detected (see Section 6.2.2(a)). These other transcripts may be premature transcription termination products or may be specific processing or degradation products of the 1.45 kb *in vivo* transcript. Since the primed-RNA species must be fractionated on non-denaturing gels, accurate sizes of the primed-RNA cannot be determined and therefore the nature of these primed-RNA species was not investigated further.

6.2.3 Transcription from the pL Promoter

6.2.3(a) Identification of the pL Transcript

The cI maintenance repressor gene was predicted to be expressed from the pL transcript (Section 6.1) and therefore the obvious source of RNA to identify the pL transcript would be a non-induced lysogen, where the cI gene should be actively transcribed.

Total cellular RNA isolated from a lysogenic culture (186cItsp) growing at 30°C was glyoxylated, fractionated on a 1.8% agarose gel and then transferred bidirectionally to nitrocellulose membranes (Section 2.22.4). One of the membranes was probed with a B-specific probe (results presented in Section 7.4.1) and the other with a cI-specific probe. The radioactive cI-specific probe was prepared by primer extending a single-stranded M13mp7 clone of a TaqI fragment of the cI gene (sequence co-ordinates 2600-2701) (Section 2.21). After digestion with BamHI and XmnI, the 109 bp BamHI/XmnI fragment was isolated from a 5% polyacrylamide gel. The BamHI site is in the polylinker region of M13mp7, some 19 bases from the insert, and the probe therefore carries 90 bases of the cI gene. Since the TaqI clone used to prepare the primer was of the same polarity as pL mRNA, hybridisation with the cI-specific probe would automatically indicate leftward transcription.

Figure 7.3(b) records the detection of a predominant 2.2 kb transcript and three minor transcripts sized at 1.3 kb, 1.1 kb and 0.85 kb in the lysogenic RNA sample, (Fig. 7.3(b), track 2). The minor transcripts may represent transcription from a weaker promoter in the vicinity of pL, hybridisation of the probe to premature transcription termination products or to processed or degraded products of the 2.2 kb transcript. Evidence is presented in Section 6.2.3(b) that all the transcripts detected with the cI-specific probe have a common 5'-end, which suggests that transcription initiates

from a single promoter. A faint band was detected in the non-lysogen control track (Fig. 7.3(b), track 1) which appeared to comigrate with the 1.3 kb band but the nature of this faint band was not investigated further.

The predominant 2.2 kb detected in the lysogenic RNA sample was consistent in size with the 2.2 kb pL transcript predicted by the location of pL and tL (Section 6.1).

6.2.3(b) Location of the 5'-end of the pL Transcript

Several attempts to confirm the *in vitro* activity of the pL promoter by sequencing the 5'-end of the associated pL transcript, using the method described in Section 6.2.1(b) and 6.2.2(b), were unsuccessful (data not shown). The problem encountered was that after annealing the cI-specific primer (109 bp BamHI-XmnI fragment used for Northern analysis in Section 6.2.3(a)) to RNA isolated from a 186cItsp lysogen, the primed-RNA species could not be detected on a 5% non-denaturing gel, even after long exposure times. The pL transcript should be expressed from only a single copy of the 186 genome in the lysogen and the pL promoter is predicted to be a weaker promoter than promoters pB and pL (Section 4.2.3); for these two reasons the amount of pL transcript in the lysogen was expected to be low and the difficulty in detecting the primed-RNA species was not unexpected.

An alternative method was used to confirm the *in vivo* activity of the pL promoter which was based on the method described by McKnight *et al.*, (1981). In this method the primer is annealed to total cellular RNA and the extension reaction carried out with AMV reverse transcriptase in the presence of all four dNTP's. Extension to the 5'-end of the RNA generates a specific primer extension product. Accurate sizing of the extension product can be achieved by comparison of its size to a DNA sequencing ladder on a denaturing 5% polyacrylamide

gel (Section 2.18.4(b). The technique is more sensitive than determining the DNA sequence from the radioactive primer to the 5'-end of the RNA transcript (used to identify the 5'-ends of the pB and pR transcripts Section 6.2.1(b), 6.2.2(b)), since the radioactive primer is extended into a single product rather than generating a DNA sequence ladder. This technique does not require fractionation and isolation of the primed-RNA species, which generally resulted in considerable losses of primed-RNA, especially when large primed-RNA species, (> 1 kb) were eluted from the gel slices (data not shown). For the reasons outlined above, the amount of RNA required to determine the 5'-end of a transcript using primer extension analysis was substantially less than that required for sequence analysis.

The radioactive 109 bp BamHI-XmnI cI-specific probe described in Section 6.2.3(a), was used as a radioactive primer on total cellular RNA isolated from lysogenic (E252) and non-lysogenic (E251) cultures growing at 30°C. Annealing was carried out by heating the radioactive primer and RNA to 100°C for 3 minutes and incubating the mixture for 90 minutes at 60°C. After annealing, the primer was extended on the RNA template in the presence of all four dNTP's and AMV reverse transcriptase as described in Section 2.22.7.

Extension products generated as a result of primer extension on lysogenic RNA were a doublet of 227 b and 230 b, (Fig. 7.4, track 4). Two smaller extension products (a doublet of 130 b and 131 b) were also detected but were present both in the lysogenic and non-lysogenic RNA tracks (Fig. 7.4, tracks 2 and 4) and were therefore not extension products on prophage mRNA. The possible nature of these extension products will be discussed in Section 7.4.2.

The number of bases extended from the 3'-end of the primer (sequence position 2688, size = 109 b) were 118 b

(227 b - 109 b) and 121 b (230 b - 109 b), placing the 5'-end of the transcript at sequence position 2806 (2688 + 118) and 2809 b (2688 + 121). Hawley and McClure (1983) used 88 promoters for which the 5'-end of the RNA transcript had been precisely determined and showed that initiation with a purine was highly preferred (82 out of 88) and that for most promoters (45 out of 82), transcription began with an A residue. For this reason the A residues in the r-strand corresponding to sequence positions 2808 and 2811 (2 bases downstream of the predicted startsites determined by primer extension) are predicted to represent the transcription startsites. The location of these predicted startsites at sequence positions 2808 and 2811 is consistent with transcription initiated from the pL promoter (-10 region of pL is located at sequence co-ordinates 2817-2822). Primer extension analysis therefore provides evidence that the promoter pL, predicted by computer analysis in Section 4.2.3, is an active in vivo promoter.

Northern analysis using the 109 b BamHI-XmnI cI-specific probe resulted in the detection of a predominant 2.2 kb lysogenic transcript (Section 6.2.3(a)) and minor transcripts of 1.3 kb, 1.1 kb and 0.85 kb. The same radioactive fragment, when used as a primer on in vivo mRNA and extended with AMV reverse transcriptase in the presence of all four dNTP's, gave only a single extension product (a doublet of 227 b and 230 b) and although not conclusive, suggests that the transcripts detected by Northern analysis have a common 5'-end which is consistent with initiation from the pL promoter.

To summarise the results of this section; evidence has been presented that the computer-predicted promoters pB, pR and pL are active promoters in vivo, by locating the 5'-ends of the associated RNA transcripts. Northern analysis was used to identify RNA transcripts

consistent in size with those predicted to initiate from the pB, pR and pL promoters. The B gene was also shown to be transcribed from a larger transcript, late in infection, as predicted from the computer analysis of the DNA sequence (Section 6.1).

SECTION 7

REGULATORY SITES

7. REGULATORY SITES

Two sites involved in the regulation of phage 186 transcription are located on the DNA sequence of the PstI(65.5%-77.4%) fragment. These two sites are the cI repressor binding site (operator site) and the 186 B gene promoter-regulatory region. Evidence for the existence of these two sites, and the role these sites play in regulating transcription, will be discussed. The rationale for the experimental approach to further define these regulatory sites is also presented.

7.1 THE OPERATOR SITE: INTRODUCTION

An operator region can be defined by the largest DNA fragment protected from nuclease digestion by bound repressor (Pirotta, 1973), or by the DNA sequence location of virulent mutations that make the phage insensitive to the immunity of the lysogenic host. These mutations represent potential changes to the repressor binding site (Maniatis et al., 1975). Three virulent mutants of phage 186 were available, and therefore the latter of the two approaches was chosen.

Virulent mutants vir1 and vir2 had been used for genetic studies of phage 186 (Hocking and Egan, 1982a) and vir3 was isolated by J.B. Egan (unpublished). Mutants vir2 and vir3 were isolated as spontaneous mutants of 186cItsp strains, whilst vir1 was a spontaneous mutant of a cIam53 strain (Section 2.2.1). Genetic studies placed vir2 to the right of the att-int region (Hocking and Egan, 1982a), but the accurate position of the three virulent mutants was not known. The following evidence was used to determine which 186 DNA fragment would be sequenced.

Hybridisation studies of Finnegan and Egan (1981) provided evidence that transcription of the early genes initiated at approximately 75% of the 186 genome and that this transcription was under cI

repressor control. One of the four major *in vitro* transcripts was shown to initiate at 74.7% (sequence position 2747) and was named the *p*R transcript (Pritchard 1984; Pritchard and Egan, 1985). Activity of the *p*R promoter was confirmed *in vivo* by isolating an RNA transcript that initiated from the same sequence position (Section 6.2.2(b)). The *p*R transcript was consequently predicted to be the early lytic transcript and the *p*R promoter should therefore be under direct cI repressor control, with at least part of the operator region represented in the DNA sequence of the *p*R promoter region.
A 258 bp *Taq*I fragment (sequence co-ordinates 2700-2959) conveniently spans the intergenic region between *c*I and *CP75*, which contains the *p*R (and *p*L) promoter region, and was therefore the fragment of choice for locating the position of the virulent mutations by DNA sequencing.

7.2 RESULTS AND DISCUSSION

DNA from three virulent mutants was digested with *Xho*I and *Bgl*II and the 3.6 kb *Xho*I-*Bgl*II(67.6%-79.6%) fragment was isolated from an agarose gel (Section 2.18.3). This fragment was sub-digested with *Taq*I and the 258 bp *Taq*I fragment, isolated after fractionation on a 5% polyacrylamide gel, was cloned into the *Acc*I site of M13mp9 (Sections 2.18.4(d), 2.19). Recombinant phage were isolated and the DNA sequence of the insert determined (Sections 2.19.7, 2.20). The 258 bp *Taq*I fragment from the *c*I*ts*p parent of *vir2* and *vir3* was also isolated, cloned and sequenced as described above. DNA sequence comparison of 186c^+ DNA (I. Lamont, personal communication) with 186*c*I*ts*p DNA (Fig. 4.5) allowed the identification of the *c*I*ts*p mutation as a point mutation, a AT-GC transition at sequence position 2634, and therefore this mutation does not lie within the DNA sequence of the 258 bp *Taq*I fragment. The sequence of the 258 bp *Taq*I fragment from the *c*I*am53* parent phage of *vir1* was not

determined.

The DNA sequences of the 258 bp TaqI fragment from each of the three vir mutants and the 186cItsp DNA, and a summary of the base-pair changes detected, are presented in Figures 7.1 and 7.2 respectively. In each case the orientation of the clone was the same and the DNA sequence generated was that of the r-strand.

All three virulent mutants had multiple base-pair changes, which was an unexpected finding since all were isolated as spontaneous mutants in 186 phage stocks. Each of the three mutants had a CG-AT transversion at sequence position 2722 and vir2 contained a further AT-GC transition at sequence position 2726. Furthermore, each of the mutants had a base-pair change at sequence position 2709, vir2 and vir3 had a CG-TA transition at this position, whilst vir1 contained a CG-AT transversion. Although the DNA sequence of 258 bp TaqI fragment of the cIam53 parent phage of vir1 was not determined, the fact that base-pair changes at sequence positions 2709 and 2722 corresponded to those found in the vir2 and vir3 mutants, which were isolated from a cItsp parent, suggests the base-pair changes at sequence positions 2709 and 2722 were important for conferring virulence in vir1.

Base-pair changes at sequence positions 2722 and 2726 were between the -35 (2712-2717) and -10 (2735-2740) hexamers of the pR promoter, confirming the prediction that the pR promoter would represent at least part of the operator region. Scanning the DNA sequence of the pR promoter region revealed three direct repeats (OD1, OD2, OD3) and three inverted repeats (OI1, OI2, OI3) (Section 4.2.5, 4.2.6), giving two sets of candidates for a tripartite operator in phage 186. OI1 (2715-2738) and OD1 (2724-2748) both overlap the -10 hexamer of the pR promoter (OI1 also overlaps the -35 hexamer), but none of the base-pair changes in virulent mutants vir1 and vir3

Fig. 7.1. DNA sequence of virulent mutants vir1, vir2 and vir3

DNA (10 ug) from mutants vir1, vir2, vir3 (Section 2.2.1), and 186cItsp DNA, was doubly-digested with XhoI and BglII, and fractionated on a 1% horizontal gel in TAE (Section 2.18.1). The 3.6 kb XhoI-BglII (67.6%-79.6%) fragment was isolated (Section 2.18.3), sub-digested with TaqI and the 258 bp TaqI fragment was isolated after fractionation on a 5% polyacrylamide gel in TBE (Section 2.18.4(a),(d)). The 258 bp TaqI fragment was cloned into the AccI site of M13mp9 and sequenced (Section 2.19, 2.20). Electrophoresis and autoradiography conditions were described in Sections 2.18.4(b) and 2.18.5 respectively.

The DNA sequence of the 258 bp TaqI fragment from (a) 186cItsp (b) vir1 (c) vir2 (d) vir3 DNA is presented. The position of base-pair changes is indicated on the autoradiograph and summarised in Fig. 7.2.

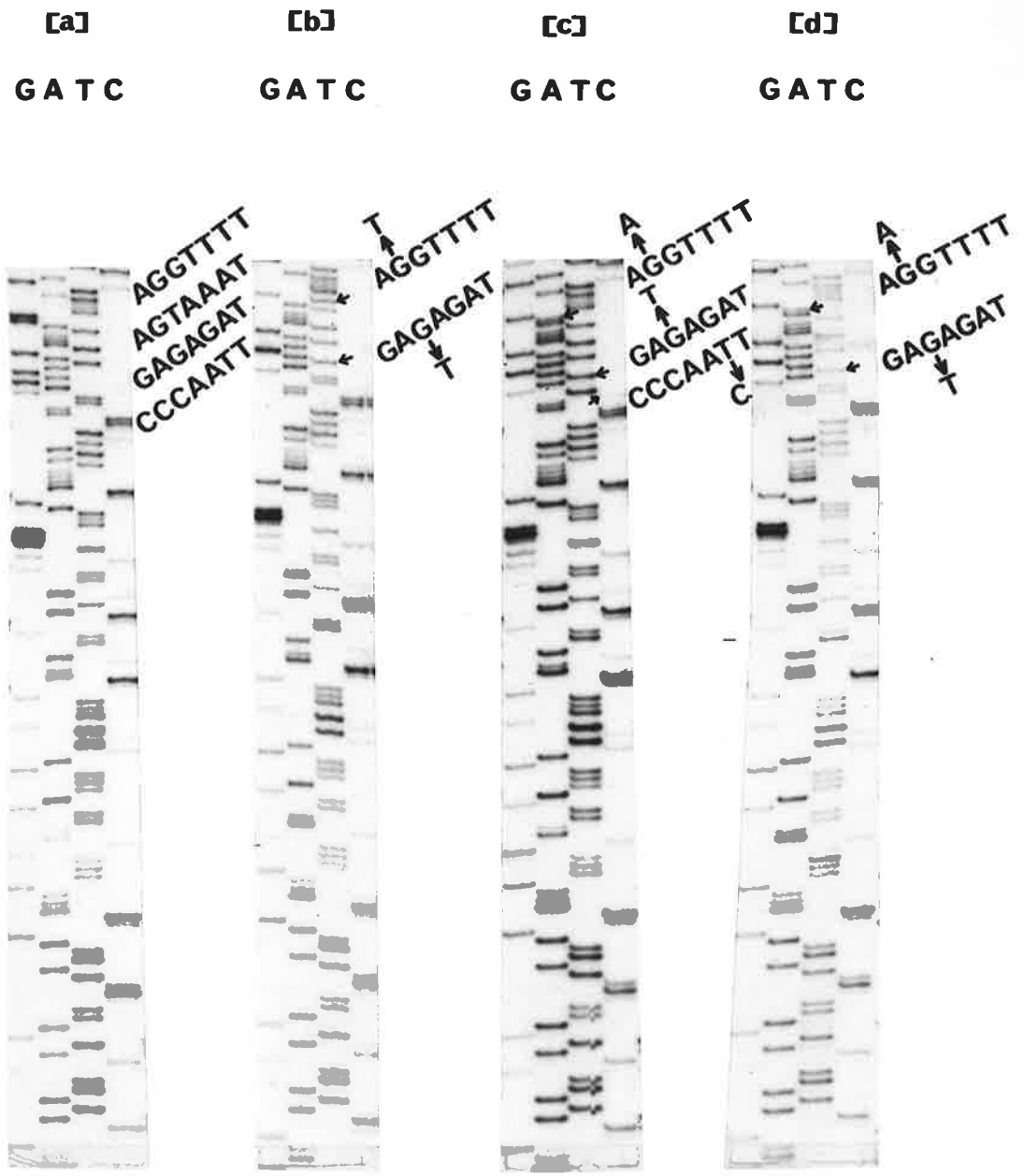
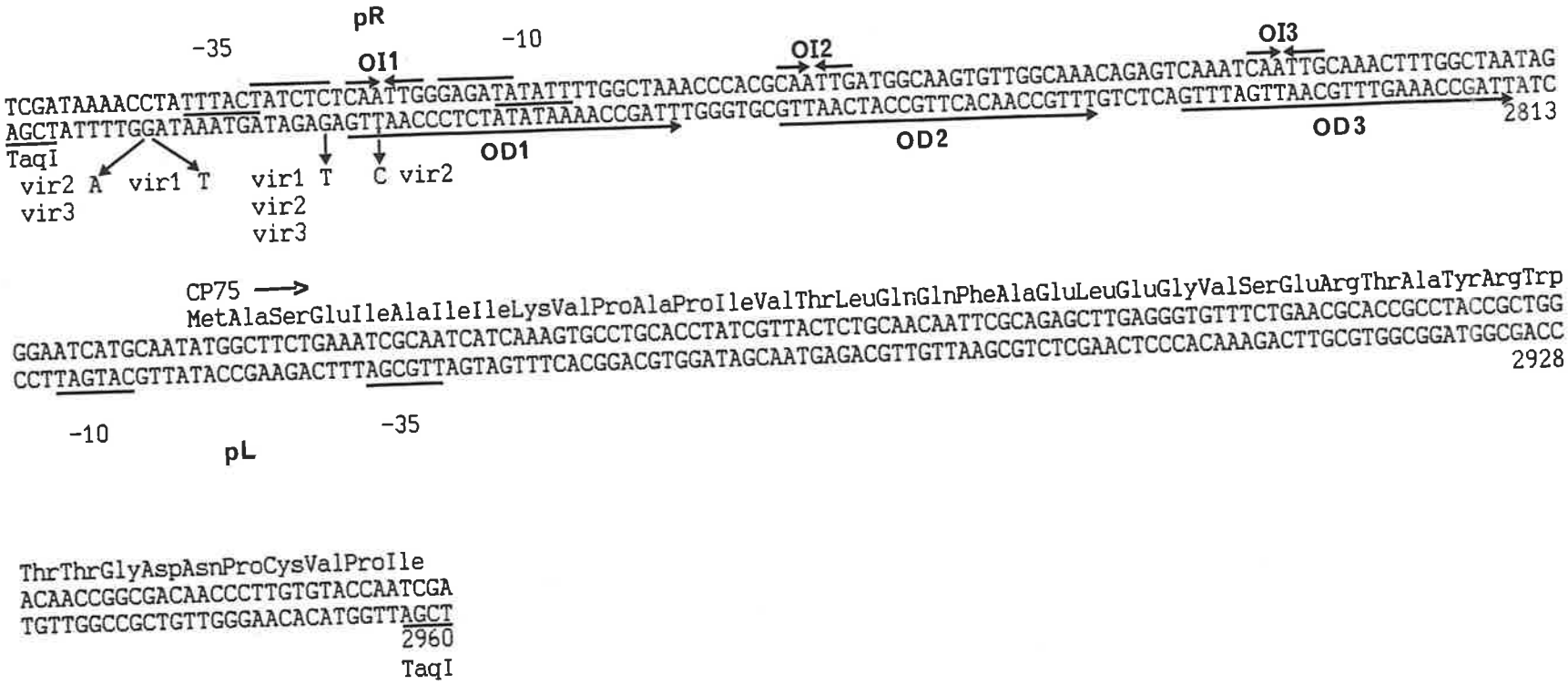

[a]
[b]
[c]
[d]
G A T C
G A T C
G A T C
G A T C
AGGTTTT
AGTAAAT
GAGAGAT
CCCAATT
T
AGGTTTT
GAGAGAT
T
A
AGGTTTT
T
GAGAGAT
CCCAATT
C
A
AGGTTTT
GAGAGAT
T

Fig. 7.2. Location of the base-pair changes detected in virulent mutants vir1, vir2 and vir3

The DNA sequence of the 258 bp TaqI fragment of 186cItsp DNA is presented and the positions of the base-pair changes detected in virulent mutants vir1, vir2 and vir3 are indicated. Also shown on the figure are the positions of the predicted pL and pR promoters (the -10 and -35 regions are underlined), the direct repeats OD1, OD2, OD3 (arrows) and inverted repeats OI1, OI2 and OI3 (head to head arrows).


pR
-35
OI1
-10
OI2
OI3
TCGATAAAACCTATTTACTATCTCTCAATTGGGAGATATATTTTGGCTAAACCCACGCAATTGATGGCAAGTGTTGGCAAACAGAGTCAAATCAATTGCAAACTTTGGCTAATAG
AGCTATTTTGGATAAATGATAGAGAGTTAACCCTCTATATAAAACCGATTTGGGTGCGTTAACTACCGTTCACAACCGTTTGTCTCAGTTTAGTTAACGTTTGAAACCGATTATC
2813
OD1
OD2
OD3
TaqI
vir2 A
vir1 T
vir1 T
C vir2
vir3
vir2
vir3
CP75
MetAlaSerGluIleAlaIleIleLysValProAlaProIleValThrLeuGlnGlnPheAlaGluLeuGluGlyValSerGluArgThrAlaTyrArgTrp
GGAATCATGCAATATGGCTTCTGAAATCGCAATCATCAAAGTGCCTGCACCTATCGTTACTCTGCAACAATTCGCAGAGCTTGAGGGTGTTTCTGAACGCACCGCCTACCGCTGG
CCTTAGTACGTTATACCGAAGACTTTAGCGTTAGTAGTTTCACGGACGTGGATAGCAATGAGACGTTGTTAAGCGTCTCGAACTCCCACAAAGACTTGCGTGGCGGATGGCGACC
2928
-10
pL
-35
ThrThrGlyAspAsnProCysValProIle
ACAACCGGCGACAACCCTTGTGTACCAATCGA
TGTTGGCCGCTGTTGGGAACACATGGTTAGCT
2960
TaqI

lie within OD1, suggesting that the direct repeat sequences OD1, OD2 and OD3 do not represent the operator binding sites. The base-pair changes at sequence 2722 and 2726 lie within OI1, but the base-pair changes at sequence position 2709 are outside OI1 and alter the proposed ribosome binding site for the cI gene (AGGT, sequence co-ordinates 2707-2710). Whether this affects translation of the cI gene is not known. The apparent requirement for at least two base-pair changes to confer virulence in phage 186 differs from the situation found in lambda where single base-pair changes resulting in the virulent phenotype were confined to conserved symmetric sequences which represent the operator sites (Maniatis et al., 1975). A study was initiated by I. Lamont and D. Ulaeto in this laboratory, to isolate more virulent mutants by UV-mutagenesis of a 186c^+ parent phage. Six virulent mutants were isolated and the DNA sequence of the intergenic region between cI and CP75 was determined for each of these mutants. Each virulent mutant contained at least one base-pair change in the region 2722 to 2732, corresponding to the region between the -10 and -35 hexamers of the pR promoter (and therefore within OI1), and at least one base-pair change (in some cases a single base-pair deletion) between sequence positions 2698 to 2709. This region corresponds to the leader sequence from the proposed ribosome binding site of the cI gene and the ATG initiation codon of the cI gene (I. Lamont, personal communication).

In summary, virulent mutants have been isolated from three different 186 parent phage strains (cItsp, cIam53, c^+). The common features of the base-pair changes detected by DNA sequencing were the presence of at least two base-pair changes in each virulent mutant and the clustering of these base-pair changes into two groups. One group was located between the -10 and -35 hexamers of the pR promoter and the second group in the leader region from the proposed

cI ribosome binding site to the ATG initiation codon. The operator site will be discussed further in Section 9.1.

7.3 THE B PROMOTER-REGULATORY REGION : INTRODUCTION

As described previously (Section 6.1), transcription of 186 DNA by E. coli RNA polymerase holoenzyme in vitro yields four major in vitro transcripts and the dominant transcript (band 4) was shown to be a transcript of the B gene. Transcription of the B gene in vitro was unexpected since B gene transcription was shown by hybridisation studies to be dependent on a functional A gene (Finnegan and Egan, 1981) and was therefore expected to have an activated promoter.

The fact that the B gene was transcribed in vitro, in the absence of phage protein synthesis, suggests that in vivo the B gene should be transcribed early after infection by unmodified E. coli RNA polymerase and further suggests that the B promoter should be active in a lysogen, unless some form of repression is operative.

Repression by the cI protein in the lysogen is unlikely, since the operator region was defined by the position of the vir2 mutant and was mapped to the right of att (Hocking and Egan, 1982a). A provocative feature of the DNA sequence containing the B promoter region was the presence of two inverted repeats named BI1 and BI2 (Section 4.2.6, Table 4.4), overlapping the -35 and -10 regions of pB. These may represent protein binding sites, in which case the possibility arises that the B protein itself could be repressing B gene transcription. To test the proposal that the B gene was expressed in the lysogen and that the B protein negatively controls its own transcription, Northern analysis and primer extension analysis were performed.

Since the 186 B and P2 ogr genes show functional (See Section

8.7) and transcriptional similarities (Pritchard, 1984; Pritchard and Egan, 1985), the prediction was made that their transcriptional control mechanism would be similar. The DNA sequence of the P2 ogr promoter-regulatory region was determined and examined for the existence of conserved features in the DNA sequence such as the inverted repeats found spanning the B promoter-regulatory region (Section 4.2.6).

7.4 RESULTS AND DISCUSSION

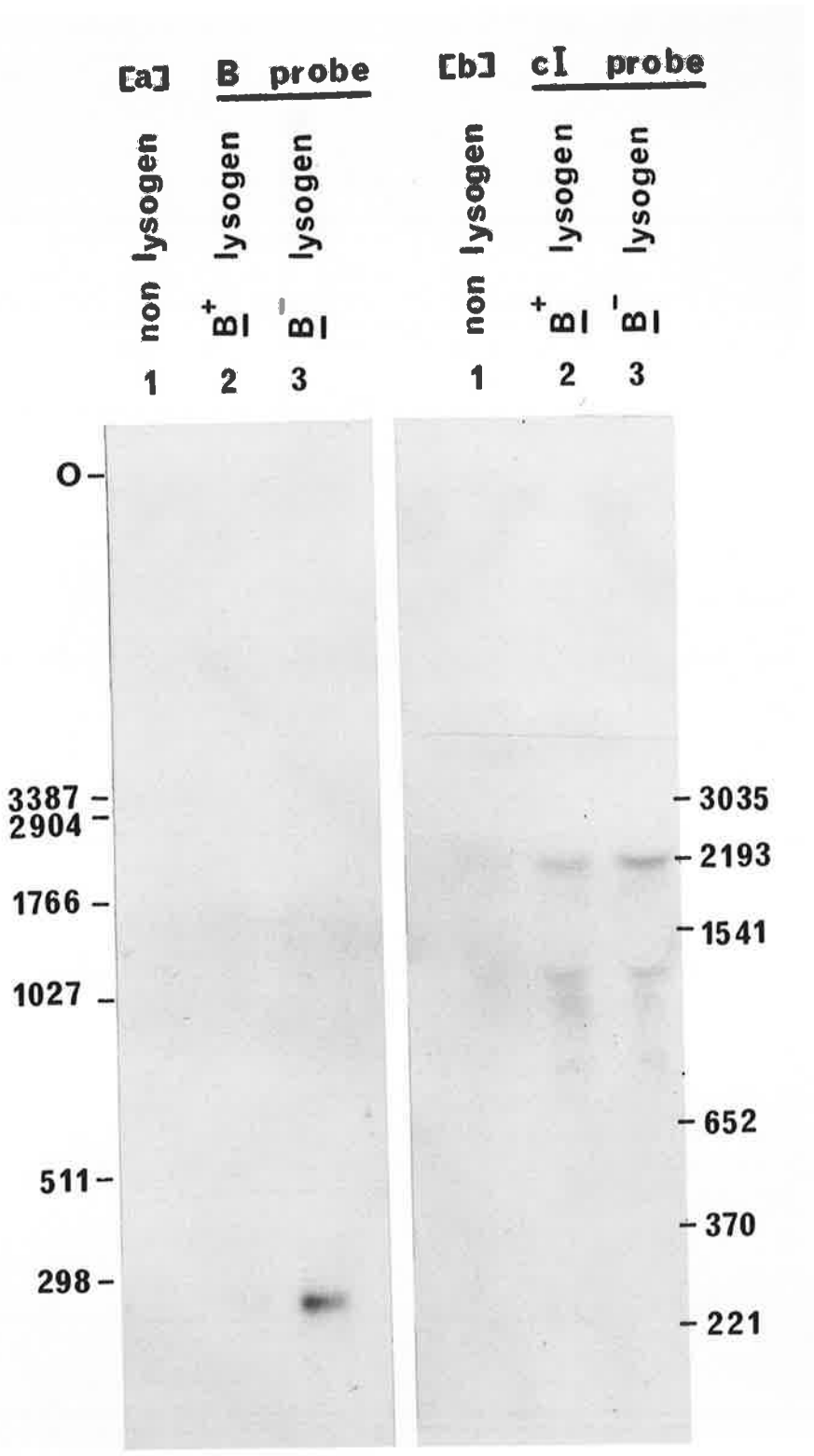
7.4.1 Northern Analysis

Total RNA was isolated from lysogenic cultures of 186cItsp growing at 30°C, glyoxylated, fractionated on a 1.8% agarose gel and transferred bidirectionally to nitrocellulose membranes (Section 2.22.4). One membrane was probed with the B-specific probe (100 bp HhaI fragment used in Section 6.2.1) and as shown in Figure 7.3(a) track 2, the 290 b B gene transcript was detected in the lysogen. Total RNA isolated from a sup° lysogen of a Bam17 mutant (referred to simply as the B^- mutant), equivalent in amount to that from the B^+ lysogen used in Figure 7.3(a) track 2, was also probed with the B-specific probe. The increased level of the B gene transcript in the B^- lysogen (Fig 7.3(a), track 3) can be readily seen, when compared to the same transcript in the B^+ lysogen (Fig. 7.3(a), track 2). As a control, the duplicate filter from the bidirectional transfer was probed with the cI-specific probe (109 bp BamHI-XmnI fragment used in Section 6.2.3.). The cI gene codes for the maintenance repressor gene which is actively transcribed in the lysogen (Finnegan and Egan, 1981) and the level of cI gene transcription should not be affected by the presence or absence of functional B protein. A predominant 2.2 kb band and minor bands at 1.3 kb, 1.1 kb and 0.85 kb (Fig 7.3 (b), tracks 2 and 3) were detected with the cI-specific probe, representing the cI maintenance

Fig. 7.3. The B gene transcript in the lysogen

RNA was isolated from a 10 ml culture growing at 30°C, as described in Section 2.22.1. 100 ug of RNA was denatured with glyoxal, fractionated on a 1.8% horizontal agarose gel and transferred bi-directionally to nitrocellulose (Section 2.22.4(a)). One membrane was probed with a B-specific probe (a) and the other with a cI-specific probe (b). The B-specific probe was a 100 bp HhaI fragment (See Section 6.2.1 and Figure 6.2(a)) and the cI-specific probe was a 109 bp BamHI/XmnI fragment (See Section 6.2.3 and Fig. 7.4). Approximately 2 x 10^6 cpm of radioactive probe was applied to each filter.

Tracks 1,2,3 in (a) and (b) carry RNA isolated from (1) non-lysogen E251, (2) B^+ lysogen E251(186cItsp) and (3) B^- lysogen E251(186cItspBam17). Glyoxylated markers were 3387 b (CMV, RNA 1), 3035 b (CMV, RNA 2), 2193 b (CMV, RNA 3), 1027 b (CMV, RNA 4), 2904 b (E. coli 23S RNA) and 1541 b (E. coli 16S RNA). Plasmid pBR325 digested with HinfI and end-labelled gave 1766 b, 652 b, 511 b (doublet 516 b + 506 b), 370 b (triplet 396 b + 394 b + 344 b) 298 b and 221 b. The autoradiograph was exposed for 5 days at -80°C with an intensifying screen.


[a] B probe
[b] cI probe
non lysogen
lysogen B+ I
lysogen B- I
1
2
3
0
3387
2904
1766
1027
511
298
3035
2193
1541
652
370
221

repressor transcripts in the lysogen. The intensities of these bands did not differ markedly whether the RNA was isolated from the $\underline{B}^+$ or $\underline{B}^-$ lysogens (Fig. 7.3 (b), track 2 and track 3).

7.4.2 Primer Extension Analysis

Primer extension analysis was also used to assess the level of the B and cI transcripts in the $\underline{B}^+$ and $\underline{B}^-$ lysogens. The B and cI-specific probes (Section 7.4.1) were used as radioactive primers and annealed in solution (after heat denaturation) to total cellular RNA from the same cultures used in Section 7.4.1. After annealing, the primer was extended on the RNA template in the presence of all four dNTP's and AMV reverse transcriptase (Section 2.22.7).

The 5'-end of the B gene transcript was predicted to be at sequence position 270 (Section 6.2.1(b)) and therefore the 100 b B-specific primer (sequence co-ordinates 334-434) should be extended 64 b on the B mRNA to give an extension product of 164 b. Primer extension of the B-specific probe yielded an extension product sized at 167b (Fig. 7.4, tracks 3 and 5) by comparison with a DNA sequencing ladder.

This extension product was not detected in the non-lysogen (Fig 7.4, track 1) and therefore represents the extension product of the B-specific primer on B mRNA present in the lysogen. The marked increase in B gene transcription, reflected by the increase in the intensity of the 167 b extension product (Fig. 7.4 track 5) could again be readily seen when the intensity of this band was compared to the same extension product in the $\underline{B}^+$ lysogen (Fig. 7.4 track 3). Extension products from the cI-specific primer were detected as a doublet, (227b and 230b, Fig. 7.4 tracks 4 and 6, see also Section 6.2.3(b)) but as expected, the intensity of this doublet showed no increase whether the extension reaction was carried out on RNA isolated from the $\underline{B}^+$ (Fig 7.4 track 4) or $\underline{B}^-$ mutants (Fig. 7.4, track 6).

Fig. 7.4. Primer extension of the B-specific and cI-specific primers on lysogenic RNA

RNA was isolated from lysogenic cultures 186cItsp (B^+), 186cItspBam17 (B^-) and a non-lysogenic culture (E251) growing at 30°C, using the method described in Section 2.22.1. The radioactive B-specific primer (a 100 b HhaI fragment described in Section 6.2.1(a)) and cI-specific primer (a 109 b BamHI/XmnI fragment described in Section 6.2.3(a)) were each added (2×10^5 cpm) to 5 ug of RNA (non-lysogen, B^+, B^-) and heated to 100°C for 3 min. Annealing was allowed to occur for 90 min at 60°C. The primer was extended on the RNA template in the presence of all four NTP's and AMV reverse transcriptase as described in Section 2.22.7.

Tracks 1,3 and 5. Primer extension of the B-specific primer on non-lysogen, B^+ and B^- RNA respectively.

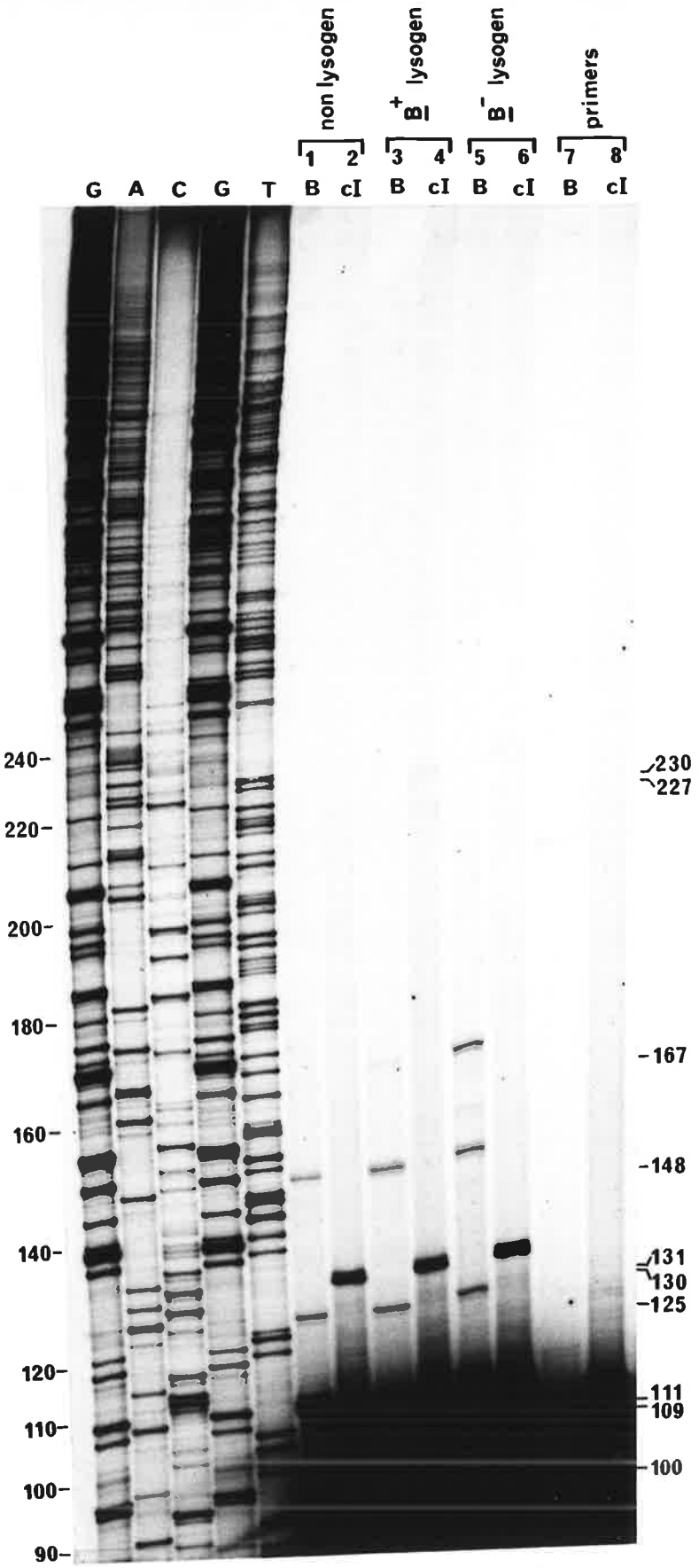
Tracks 2,4 and 6. Primer extension of the cI-specific primer on non-lysogen, B^+ and B^- RNA respectively.

Track 7 contains B^+ RNA and B-specific primer which were treated as described for Tracks 1-6 except that AMV reverse transcriptase was omitted from the reaction mix.

Track 8 is identical to Track 7 except the cI-specific primer was used.

A sequencing ladder was included to provide size markers and the size given on the left of the figure is from the 5'-end of the M13 universal primer. The gel shows some distortion of Tracks 5-8 and all sizes of extension products were therefore estimated from Tracks 1-4 and are shown on the right of the figure.

Electrophoresis conditions were given in Section 2.18.4(b) and the autoradiograph was exposed for 24 hours at -80°C with an intensifying screen.


non lysogen
B+ lysogen
B- lysogen
primers
1 2 3 4 5 6 7 8
G A C G T B cI B cI B cI B cI
240
220
200
180
160
140
120
110
100
90
230
227
167
148
131
130
125
111
109
100

In conclusion, Northern analysis (Section 7.4.1) and primer extension analysis (Section 7.4.2) have provided evidence that the late control gene B is indeed transcribed in the lysogen and that B gene transcription appears to be negatively controlled by the B protein itself (negative autogenous control). However the possibility that the control of B gene transcription is through a protein under B gene control in the lysogen, and therefore indirectly controlled, cannot be totally excluded.

Several extension products were in fact produced from both primers but all were present for both lysogenic and non-lysogenic RNA. These products were a result of the primer extension reaction since they disappeared in the controls where extension was prevented (Fig. 7.4 tracks 7 and 8). A likely possibility is that the extension products found for both lysogenic and non-lysogenic RNA were generated by "loop-back" of the primer on itself which resulted in self-priming. This is a problem sometimes encountered when large primers (>50bp) are used for primer extension analysis (A. Kultonow, personal communication). A control track which should therefore have been included was to carry out the primer extension reaction on the primer itself without any RNA in the reaction mix.

7.4.3 Determination Of The P2 ogr Promoter-Regulatory Region

The source of DNA for sequencing was a 3.0 kb PstI-BglII (62.6%-77.2%) fragment from P2vir22ogr52 DNA (Section 2.2.2) cloned into M13mp9 (supplied by M. Pritchard). The 5'-end of the 290 b P2 ogr in vitro RNA transcript was sequenced using a 69 bp radioactive HhaI restriction fragment as a primer (Pritchard, 1984; Pritchard and Egan, 1985). This same radioactive fragment was used to determine the DNA sequence of the P2 ogr promoter-regulatory region.

The M13mp9 clone of the PstI-BglII(62.6%-77.2%) fragment

of P2 vir22ogr52 DNA was used as a template to synthesise a radio-labelled copy of the insert and, after HhaI digestion, the radioactive 69 bp HhaI fragment (sequence co-ordinates 304-373, Fig. 7.5(b)) was isolated (Section 2.21). This radioactive fragment was used as a primer by annealing it back to an M13mp9 clone of the PstI-BglII (62.6%-77.2%) fragment of P2vir22ogr52 DNA and the DNA sequence from the radioactive primer toward the P2 ogr promoter-regulatory region was determined as described in Section 2.20.3.

To confirm the DNA sequence of the P2 ogr promoter-regulatory region from both DNA strands, a 0.5 kb BglI fragment, predicted from the DNA sequence gained by primer extension to span the P2 ogr promoter-regulatory, region was used for subcloning. The 0.5 kb BglI fragment was isolated after BglI digestion of the 3.0 kb PstI-BglII fragment of P2vir22ogr52 DNA and fractionation on a 1% agarose gel (Section 2.18.1, 2.18.3). This was digested with HpaII and the resulting fragments were fractionated on a 5% polyacrylamide gel (Section 2.18.4(a)). Three HpaII fragments (29 bp, 219 bp, 171 bp) were isolated and cloned into the AccI site of M13mp9 (Section 2.18.4(d), 2.19). M13 clones representing both strands of the three HpaII fragments were sequenced.

The sequencing strategy and DNA sequence of the P2 ogr promoter-regulatory region are presented in Figure 7.5(a) and (b) respectively. Transcription of ogr mRNA was predicted to initiate at the A residue at sequence position 239 (Pritchard, 1984; Pritchard and Egan, 1985) and associated with this transcript was a promoter sequence, pogr, with a -10 region, TAAAAT (sequence co-ordinates 227-232), and a -35 region, TTGTTT (sequence co-ordinates 204-209). This promoter was detected by the SCAN program using the promoter weight matrix (Section 2.23.2) and has a promoter homology score (promoter strength) of 1.9. The pogr promoter, on the basis of

Fig. 7.5. Cont'd

(b) DNA sequence of the P2 promoter-regulatory region. Reading frames predicted to be those of the P2 D and ogr genes are indicated. The position of the ogr52 mutation, determined by Pritchard (1984) is shown. The sequence to the right of the NruI site (O) was independantly confirmed by E. Ljungquist (personal communication). The -10 and -35 regions of the pogr promoter, as predicted by the SCAN program (Section 2.23.2), are underlined. Also indicated is the predicted transcription start point of the ogr mRNA (Pritchard, 1984; Pritchard and Egan, 1985) and the predicted ribosome binding site (RBS). Numbering of the sequence is arbitrary.

Fig. 7.5. DNA sequence of the P2 ogr promoter-regulatory region

(a) Features of the P2 ogr promoter-regulatory region as predicted by sequence analysis (Pritchard, 1984; E. Ljungquist, personal communication; this work, Section 7.4.3) and by primer extension of the 69 b radioactive HhaI primer (sequence positions 304-373, Fig. 7.5(b)) on an M13mp9 clone of the 3.0 kb PstI-BglII(62.6%-77.2%) fragment from P2vir22ogr52 DNA. The position of the pogr promoter and togr terminator are shown relative to the ogr gene and the vir22 deletion at 72.2% (See Section 2.2.2). Also indicated in the position of the predicted P2 tail gene D.

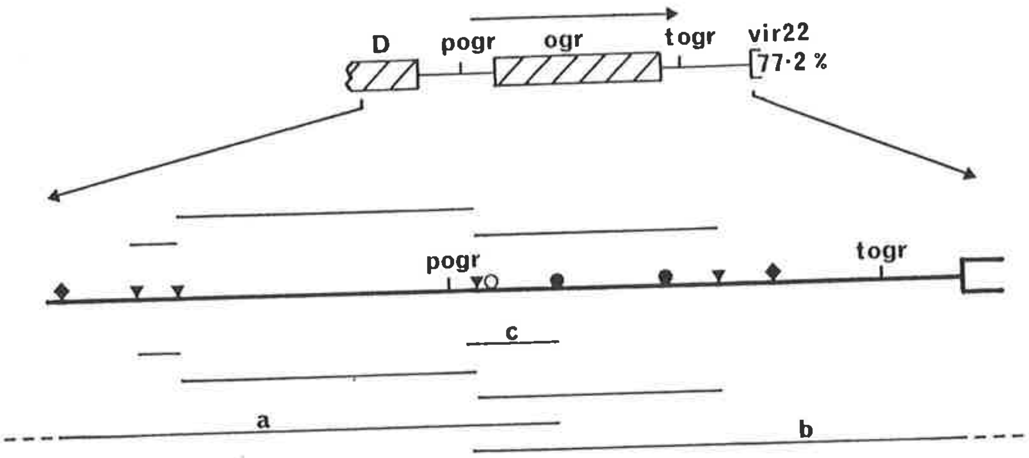
Expanded below the figure is the sequencing strategy for the P2 promoter-regulatory region. Restriction sites used for sub-cloning and sequencing of the 0.5 kb BglI fragment are indicated BglI(◆), HpaII(▼), HhaI(●). Horizontal bars indicate the extent of the DNA sequence determined, those above the line were sequences generated of the l-strand and those below were sequences were generated of the r-strand.

a. The 69 b HhaI radioactive primer, generated by primer extension of an M13mp9 clone of the 3.0 kb PstI-BglII(62.6%-77.2%) (Sections 7.4.3, 2.21), was annealed back to an M13mp9 clone of the same 3.0 kb fragment as described in Sections 7.4.3 and 2.20.3. The DNA sequence from the radioactive primer was determined as described in Section 2.20.3.

b. DNA sequence determined from the BglII site(77.2%) (not shown) of the 3.0 kb PstI-BglII(62.6%-77.2%) fragment of P2vir22ogr52 DNA, which was cloned into M13mp9 (Pritchard, 1984).

c. Sequence determined by primer extension of the 69 b HhaI radioactive primer (See (a)) on the P2 ogr in vitro RNA transcript. (Pritchard, 1984; Pritchard and Egan, 1985).

[a]



[b]

```
D--->
 Gly Arg Ala Asp Ile Tyr Thr Glu Thr Pro Val Lys Val Ser Gly Phe Lys Arg Val Ile Asp Glu Gln Asp Trp Thr Ile Thr Lys Val
 GGC AGG GCA GAT ATT TAC ACG GAA ACA CCG GTC AAA GTG TCT GGC TTT AAG CGC GTC ATA GAC GAG CAG GAC TGG ACA ATC ACT AAG GTG
CCGGT CGG GCA GAT ATT TAC ACG GAA ACA CCG GTC AAA GTG TCT GGC TTT AAG CGC GTC ATA GAC GAG CAG GAC TGG ACA ATC ACT AAG GTG
HpaII                                HpaII                                                                             92

 Thr His Phe Leu Asn Asn Ser Gly Phe Thr Thr Ser Leu Glu Leu Glu Val Arg Leu Ser Asp Val Glu Tyr Glu Thr Glu Asp Asp Glu
 ACA CAT TTT CTG AAT AAT AGC GGC TTC ACG ACG TCC TTA GAG CTT GAG GTC AGG CTT TCT GAT GTG GAG TAC GAA ACA GAA GAT GAT GAG
                                                                                                                       182

                                                                                     ogr --->
                                                                                     Met Phe His Cys Pro Leu Cys Gln His
***
TGATGTTTTTGTTTTATCTGTTTGTTTTGTAAGGATAAATTAACTAAAATGGCACCATCAACAAAACCGGAAGAGGTGCTCGCGATG TTT CAT TGT CCT TTA TGC CAG CAT
                      -35      pogr          -10      +1          HpaII RBS   NruI                                 293

 Ala Ala His Ala Arg Thr Ser Arg Tyr Ile Thr Asp Thr Thr Lys Glu Arg Tyr His Gln Cys Gln Asn Val Asn Cys Ser Ala Thr Phe
 GCC GCA CAT GCG CGT ACA AGT CGC TAT ATC ACT GAC ACG ACA AAA GAG CGT TAT CAT CAG TGC CAG AAC GTG AAT TGC AGC GCC ACG TTC
             HhaI                                                                                     HhaI         383

 Ile Thr Tyr Glu Ser Val Gln Arg Tyr Ile Val Lys Pro
 ATC ACT TGT GAG TCG GTA CAG CGA TAC ATC GTG AAG CCGG
          A ogr+                                 HpaII(423)
```

promoter homology scores, is predicted to be weaker than the 186 pB promoter (promoter strength = 2.4) but stronger than the 186pR promoter (promotor strength = 1.2).

The reading frame commencing with the ATG codon at sequence positions 267 to 269 was preceded by a potential ribosome binding site (GGAG, sequence co-ordinates 255-258) which conformed to Stormo rule 6 (Stormo et al., 1982). Pritchard (1984) identified this reading frame as that of the ogr gene by determining the DNA sequence of the ogr52 mutation, which was shown to be an AT-GC transition at sequence position 391, as indicated on Figure 7.5(b). The reading frame was shown to be 72 amino acids long (Pritchard, 1984; Pritchard and Egan, 1985) and shows extensive amino acid sequence homology to the B protein (Section 8.7). A second open reading frame commenced to the left of the first HpaII site on the DNA sequence in Figure 7.5(b) and terminated at the TGA stop codon at sequence positions 183 to 185. This reading frame is predicted to be that of the P2D gene, which maps immediately to the left of the ogr gene (Bertani et al., 1982). This reading frame shows extensive amino acid homology to the 186 D protein (Section 8.7).

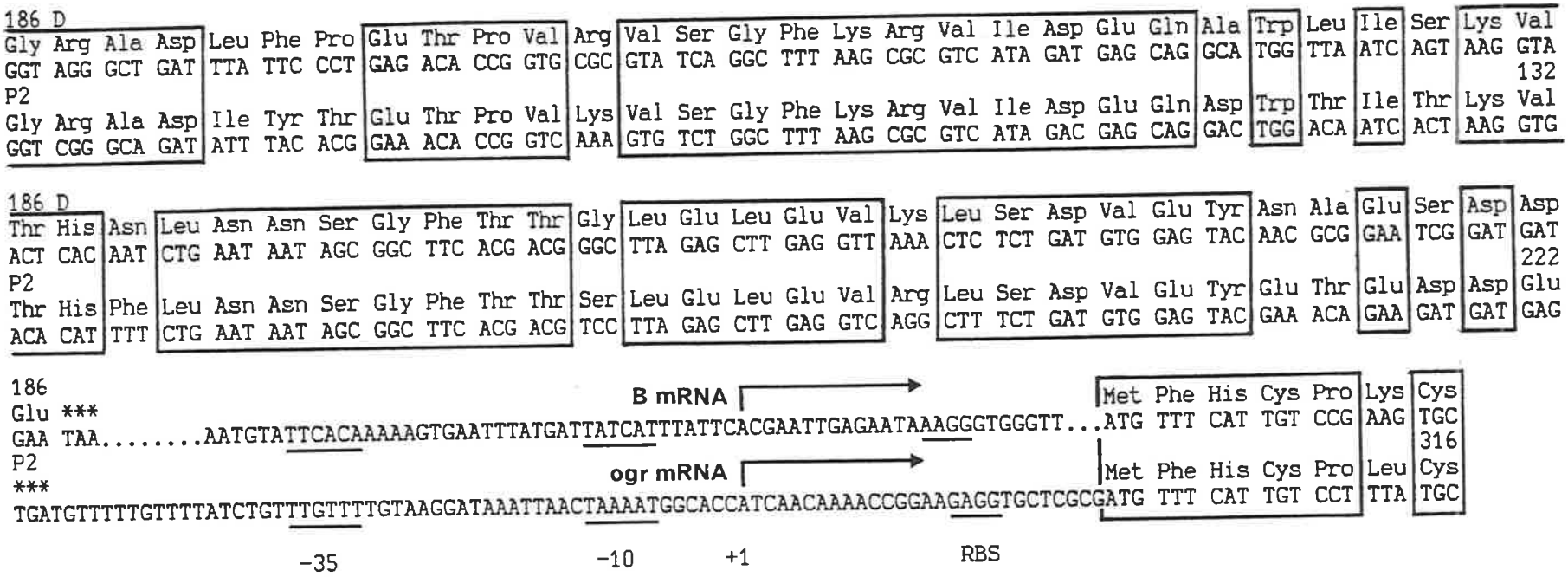
A comparison of the P2 ogr and 186 B promoter-regulatory regions is presented in Figure 7.6. The P2 D-ogr and 186 D-B intergenic regions were different in length (81 bp and 66 bp respectively) and inverted repeat sequences similar to those found spanning the -10 and -35 regions of the 186 pB promoter, were not found in the P2 D-ogr intergenic region. The absence of any significant homology between the 186 D-B and P2 D-ogr intergenic regions suggests that the control of 186 B and P2 ogr gene transcription may be quite different. Control of B and ogr gene transcription will be discussed further in Section 9.4.

Potential rho-independent terminator structures were not

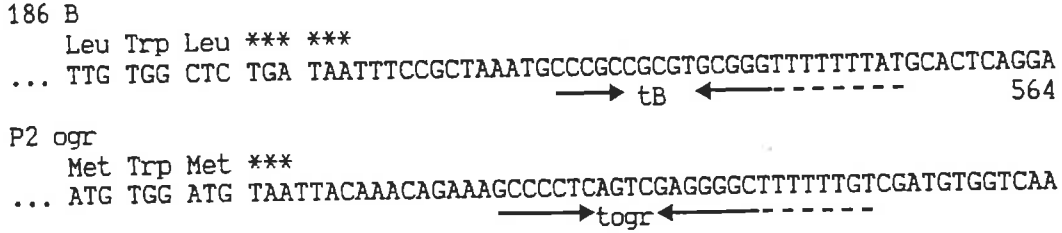
Fig. 7.6. Sequence comparison of the intergenic regions of 186 D-B and P2 D-ogr

(a) Amino acid sequence homology between the 186 D gene and the predicted P2 gene D is boxed, as is the amino-terminal homology between the 186 B and P2 ogr genes. The predicted 5'-end for the B transcript, both in vivo (Section 6.2.1(b)) and in vitro (Pritchard and Egan, 1985), is indicated. The predicted 5'-end of the P2 ogr transcript is also indicated (Pritchard and Egan, 1985). Promoter regions are aligned at the -10 and -35 regions and the proposed ribosome binding sites (RBS) are also shown.
(b) Proposed rho-independent terminator regions of the 186 B and P2 ogr transcripts (tB and togr). Inverted repeats are indicated by the head to head arrows and the consecutive T residues which correspond to the poly U rich region in the mRNA are dashed. Both terminators fit the criteria for a rho-independent terminator as described by Rosenberg and Court (1979).

[a] Promoters



[b] Terminators



found in the P2 D-ogr intergenic region, implying that the P2 ogr gene may be expressed from a larger transcript late in infection, as was shown to be the case for the 186 B gene (Section 6.2.1(a)).

SECTION 8

ANALYSIS OF THE PREDICTED PROTEINS

8. ANALYSIS OF THE PREDICTED PROTEINS

8.1 INTRODUCTION

Evidence that identifies the 186 genes of the early control region, which were predicted from the DNA sequence of the PstI (65.5%-77.4%) fragment, was presented in Section 5. The aim of this Section is to describe the features of each of the predicted proteins, with particular emphasis on proteins predicted to be involved in the control of transcription (B and cI) and the integration process (Int). Proteins of phage 186 were also compared with those of related function found in the P2-related phages and in the lambdoid phages.

A brief description of the particular features searched for in each of the predicted phage 186 proteins, and the reasons why these features were of interest, is presented.

8.1.1 General Properties

General properties of the proteins translated from the DNA sequence such as the frequency of basic, acidic and hydrophobic residues and the codon usage pattern are presented in Table 8.1. The predicted molecular weight of each protein (calculated by the MWCALC program of Staden, (1980)) is also given and compared with the molecular weight of proteins expressed in the maxicell system from plasmid clones of the corresponding genes (Section 5).

8.1.2 Polarity

Capaldi and Vanderkooi (1972) calculated the polarity of a large number of soluble and membrane-associated proteins by summing the mole fraction of polar amino acids to give a polarity index. The majority of soluble proteins had a polarity index of 47%$\pm$6% whereas a considerable proportion of known membrane proteins (9 out of 19 known membrane proteins) had polarity indexes significantly lower than those of the majority of soluble proteins. A polarity

Table 8.1 Properties of Proteins Predicted from the DNA sequence

		B	CP69	Int	cI	CP75
Ala	GCA	2	4	7	5	5
	GCC	2	2	4	2	1
	GCG	1	1	6	3	1
	GCT	0	1	7	5	2
		5	8	24	15	9
Arg	AGA *	0	1	6	4	0
	AGG *	0	0	1	2	0
	CGA *	1	1	3	2	0
	CGC	3	3	10	1	5
	CGG *	1	0	5	0	0
	CGT	0	0	4	2	3
		5	5	29	11	8
Asn	AAC	3	0	8	5	1
	AAT	0	2	8	2	0
		3	2	16	7	1
Asp	GAC	1	0	4	6	1
	GAT	0	5	10	8	0
		1	5	14	14	1
Cys	TGC	2	0	0	2	1
	TGT	3	0	5	1	1
		5	0	5	3	2
Gln	CAA	0	1	5	4	3
	CAG	3	0	5	2	1
		3	1	10	6	4
Glu	GAA	3	0	14	7	4
	GAG	1	1	14	6	3
		4	1	28	13	7
Gly	GGA *	1	8	3	5	1
	GGC	0	3	10	4	3
	GGG *	0	2	0	4	0
	GGT	2	4	8	3	3
		3	17	21	16	7
His	CAC	2	0	8	2	0
	CAT	4	1	5	1	1
		6	1	13	3	1
Ile	ATA	1	0	1	2	0
	ATC	1	0	5	5	7
	ATT	2	0	5	7	2
		4	0	11	14	9
Leu	CTA *	1	0	5	3	0
	CTC	1	1	8	3	1
	CTG	0	0	7	0	1
	CTT	0	3	3	3	1
	TTA	0	2	7	6	0
	TTG	1	0	5	4	2
		3	6	35	19	5
Lys	AAA	1	0	22	9	4
	AAG	1	0	4	2	2
		2	0	26	11	6
Met	ATG	3	2	11	2	1
		3	2	11	2	1
Phe	TTC	0	0	8	4	1
	TTT	3	0	5	5	1
		3	0	13	9	2
Pro	CCA	1	5	5	0	1
	CCC	0	1	1	0	1
	CCG	5	0	3	2	1
	CCT	1	4	4	1	3
		7	10	13	3	6
Ser	AGC	1	0	4	5	0
	AGT	1	1	3	2	0
	TCA	0	1	3	3	0
	TCC	0	0	1	4	1
	TCG	0	0	2	2	0
	TCT	0	1	2	2	2
		2	3	15	18	3
Ter (Stop)	TAA	1	1	1	1	1
	TAG	0	0	0	0	0
	TGA	1	0	0	1	0
		2	1	1	2	1
Thr	ACA	2	1	5	5	2
	ACC	1	2	8	2	2
	ACG	3	0	3	1	0
	ACT	2	3	5	0	1
		8	6	21	8	5
Trp	TGG	1	0	6	3	2
		1	0	6	3	2
Tyr	TAC	1	1	6	3	2
	TAT	1	7	5	2	1
		2	8	11	5	3
Val	GTA	0	0	3	3	1
	GTC	1	2	2	1	1
	GTG	0	0	4	4	1
	GTT	1	6	5	4	2
		4	8	14	12	5
TOTAL		74	84	337	194	88
BASIC[a]		7	5	55	22	14
BASIC (%)		9.7	6.0	16.4	11.5	16.1
ACIDIC[a]		5	6	42	27	8
HYDROPHOBIC[a]		31	51	148	93	46
POLARITY INDEX[b]		46	29	51	47	41
MODULATING CODONS *[c]		4	12	23	20	1
MODULATING CODONS * (%)		5.6	14.5	6.8	10.4	1.1
MOL. WT.[d]		8270	8405	38925	21232	9770
GEL MOL. WT.		nd[e]	8400	38900	21200	9800

Notes to Table 8.1

a. Basic (Lys+Arg), acidic (Glu+Asp) and hydrophobic (Ala+Val+Leu+Ile+Phe+Trp) residues were calculated to allow a comparison with the tabulation of these residues in lambda proteins (Daniels et al., 1982).
b. Proteins below a polarity index of 40% are considered likely candidates for membrane-associated proteins (Capaldi and Vanderkooi, 1972).
c. Proteins that are strongly expressed have a low percentage (0.6%) and weakly expressed proteins have a higher percentage (>5.2%) of modulating codons (AGA,AGG,CGA,CGG,GGA,GGG,CTA) as described by Grosjean and Fiers, (1982).
d. Estimated size of the proposed corresponding translation product in the maxicell system as described in Section 5.
e. Not determined

index of less than 40% suggests that the protein has a low polarity and is likely to be a membrane-associated protein.

8.1.3 Modulating Codons

Grosjean and Fiers (1982) provided statistical evidence that rare codons corresponding to minor tRNA species in E.coli (so-called modulating codons), regulate the elongation rate during translation; highly expressed genes on average contain a low percentage of modulating codons (0.6%), and poorly expressed genes a high percentage (5.2%).

8.1.4 DNA-binding Proteins

8.1.4(a) General

Proteins with a high frequency of basic amino acids have the potential for interaction with the negatively charged phosphate backbone of DNA and are therefore possible candidates for control proteins. The Int, Cro and cII and cIII control proteins of phage λ all have a high basic amino acid content (>15%, Daniels et al., 1982). Although λ cI repressor protein does not show a high basic amino acid content (10.6%), Sauer and Anderegg (1978) noted the basic amino acids in this protein were clustered in the amino-terminal domain (10 basic amino acids in the first 40) and speculated these residues interacted with the phosphate backbone of the DNA. Clustering of basic amino acid residues may therefore also indicate a protein with the potential to interact with the DNA phosphate backbone.

8.1.4(b) The α2-α3 DNA-binding Helix Motif of Site-specific DNA-binding Proteins

Three regulatory site-specific DNA-binding proteins have been studied by X-ray crystallography and their three-dimensional structures have been determined. Catabolite activator protein (CAP) of E. coli (McKay and Steitz, 1981), the amino-terminal region of λ cI repressor (Pabo and Lewis, 1982) and λ Cro repressor (Anderson

et al., 1981) all bind to different DNA sequences but show structural similarities in the proposed DNA-binding region, namely a pair of α helices generally referred to as the α2-α3 helices (Steitz et al., 1982; Matthews et al., 1982; Ohlendorf et al., 1982). Sequence homology with the α2-α3 helical region was detected in many other proteins that regulate gene expression (Matthews et al., 1982; R.T. Sauer et al., 1982), suggesting the α2-α3 helical DNA-binding region is a common motif in many prokaryotic site-specific DNA-binding proteins. A recent study by Laughon and Scott (1984) revealed a conserved protein domain in the ftz gene of Drosophila melanogaster and several other homoeotic genes of eukaryotes showed structural homology to the α2-α3 helical DNA-binding region, suggesting that the molecular principles derived from studying prokaryotic site-specific DNA-binding proteins may be relevant to studies of eukaryotic proteins which are proposed to act by binding to specific sites on the DNA.

Pabo and Sauer (1984) aligned the amino acid sequences of the proposed α2-α3 helical DNA-binding regions of eighteen site-specific DNA-binding proteins with those of λ cI, λ Cro and E. coli CAP proteins. An amino acid weight matrix was generated from the compilation and used in the SCAN program (Section 2.23.2, modified by I. Dodd for amino acid sequence analysis) to search for potential DNA-binding regions in the predicted phage 186 proteins.

8.2 cI PROTEIN

8.2.1 Size and Composition

Program GENE (Section 2.23.1) detected the predicted 186 cI gene reading frame when the codon usage standard of the λ cI gene was used. (Fig. 4.1). A protein of 21 Kd was predicted to be encoded by the cI gene, and a 22.7 Kd protein was identified as a potential candidate for the cI protein in the maxicell system, using the plasmid clone pEC601 (Section 5.2.3).

The cI protein is comparable in size with the 26 Kd λ cI repressor protein (Sauer and Anderegg, 1978) and both 186 and λ repressor proteins are considerably larger than the 11 Kd C repressor protein of P2 (Lundqvist and Bertani, 1984). Neither DNA nor amino acid sequence homology was detected between the cI gene of 186 and the C gene of its close relative, P2. This situation is in marked contrast to that found in several members of the lambdoid phage family (λ, 434 and P22), where significant DNA and amino acid sequence homology is found between the repressor genes (R.T. Sauer et al., 1982).

Phage 186 cI repressor (11.5% basic) and λ cI repressor (10.6% basic; Daniels et al., 1982) proteins have a higher basic amino acid content than the P2 C protein (7.1% basic; Ljungquist et al., 1984). The λ cI repressor shows a clustering of basic amino acids in the amino-terminal domain (Sauer and Anderegg, 1978), with 10 basic amino acids in the first 40 residues, whereas the 186 cI repressor and P2 C repressor do not show this clustering (4 basic amino acids each in the first 40 residues).

The 186 cI and P2 C genes have a high percentage of modulating codons (10.4% and 5.1% respectively) and would be expected to show slow translational elongation rates.

8.2.2. cI Repressor and Other Site-specific DNA-binding Proteins

A weight matrix generated from the compilation of regulatory site-specific DNA-binding proteins (Pabo and Sauer, 1984) was used in the SCAN program to search for potential DNA-binding proteins (See Section 8.1.4(b)). Matrix scores are presented in Table 8.2 and a threshold score of 76 was chosen, since it represented the lowest score obtained from the amino acid sequences in the compilation. Amino acid sequences of 186 and P2 proteins predicted to contain the α2-α3 DNA-binding helix motif are aligned with the predicted

Table 8.2 DNA-binding proteins

Protein[a]	Matrix[b] Score	Position[c]	Protein	Matrix Score	Position
λ Rep	108	41	P22 Arc[d]	65	
λ Cro	110	24	P22 Mnt[d]	57	
P22 Rep	101	29			
P22 Cro	122	21	186 Proteins[e]		
434 Rep	115	25	----------		
434 Cro	104	27	D	52	10
CAP	99	177	B	45	49
Fnr	99	195	CP69	49	36
Lac Rep	117	14	Int	62	211
Gal Rep	118	12	cI	60	34
λ cII	113	32	CP75	85	25
P22 cI	125	34	CP76	89	33
Mat ∝	101	125			
Tet Rep Tn10	117	35	P2 Proteins[f]		
Tet Rep pSC101	106	35	----------		
Trp Rep	101	74	D	57	159
H-inversion	104	170	Ogr	46	39
Tn3 Resolvase	97	169	Int	62	117
γδ Resolvase	76	169	B	58	97
Ara C	86	204	C	84	28
Lex Rep	83	36	Cox	79	22

Notes to Table 8.2

a. Protein sequences (Pabo and Sauer,1984) used to generate matrix.
b. As described in Sections 8.1.4(b) and 2.23.2.
c. Residue number corresponding to the highly conserved glycine residue.
d. Protein sequence from Sauer et al. (1983).
e. Proteins referred to in Section 8.
f. Protein sequences by conceptual translation of DNA sequences provided by E. Ljungquist (personal communication) and from Ljungquist et al. (1984).

DNA-binding helices of λ cI, λ Cro and E. coli CAP proteins in Figure 8.1. The position referred to in Table 8.2 and Figure 8.1 is that of the highly conserved glycine residue in the region between the two helices, unless otherwise specified.

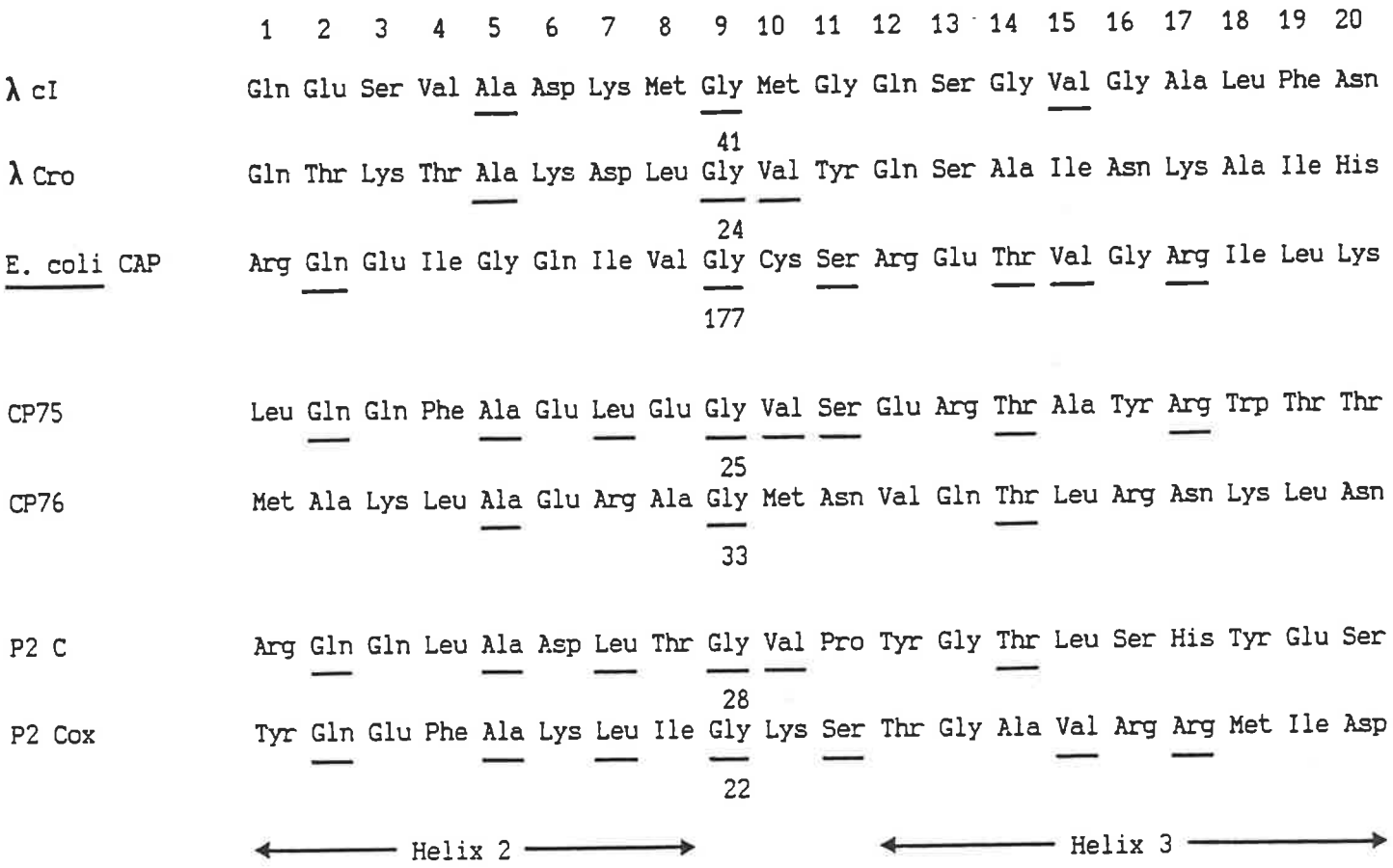
Ljungquist et al. (1984) aligned the P2 C repressor protein with six other DNA-binding proteins (including λ cI and Cro) and found the optimal alignment of the sequences was made when glycine residue number 28 of the P2 C protein was aligned with the highly conserved glycine residue between the α2 and α3 helices of λ cI and Cro. This glycine residue was successfully predicted by the weight matrix, and a score of 84 (Table 8.2, Fig. 8.1) was given for the alignment, suggesting the likely presence of the α2-α3 DNA-binding helix motif.

Predicted proteins of phage 186, CP75 and CP76, score significantly above the threshold score in a single region near the amino-terminal region of the proteins, corresponding to glycine residues 25 and 33 respectively, and must be considered candidates for site-specific DNA-binding proteins. Scores above the threshold value were not found for proteins D, B, Int and CP69 (Table 8.2). The Int protein of phage λ is a site-specific DNA-binding protein but this type of protein is expected to show a different interaction with substrate DNA from that predicted for the site-specific DNA-binding proteins which show the α2-α3 DNA-binding helix motif (Better et al., 1983).

Contrary to expectations the 186 cI repressor scored well below the threshold score of 76 (Score = 60, Table 8.2) and in this respect is similar to the Arc and Mnt proteins of P22, which presumably act as transcriptional repressors by binding to DNA (Sauer et al., 1983), but also fail to score above the threshold of detection for the α2-α3 DNA-binding helix motif. This observation

Fig. 8.1. Comparison of the amino-terminal sequence of predicted 186 and P2 proteins with three known DNA-binding proteins

The amino-terminal protein sequences are aligned as suggested by Matthews et al. (1982). Residues that are common to two or more proteins are underlined and the number of amino acids from the amino-terminus to the highly conserved glycine residue in the turn region between the two α helices is given. Residues 1-8 and 12-20 span the regions which contain the ∝2 structural helix and the ∝3 DNA recognition helix Matthews et al. (1982). Predicted 186 proteins CP75, CP76 and P2 proteins C and Cox, score significantly above the threshold of detection, using an amino acid weight matrix generated from the compilation of known site-specific DNA-binding proteins (Section 8.1.4(b)), at glycine residues 25, 33, 28 and 22 respectively. The source of the P2 C amino acid sequence was Ljungquist et al. (1984) and the P2 Cox amino acid sequence was from E. Ljungquist (personal communication).


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
λ cI Gln Glu Ser Val Ala Asp Lys Met Gly Met Gly Gln Ser Gly Val Gly Ala Leu Phe Asn
41
λ Cro Gln Thr Lys Thr Ala Lys Asp Leu Gly Val Tyr Gln Ser Ala Ile Asn Lys Ala Ile His
24
E. coli CAP Arg Gln Glu Ile Gly Gln Ile Val Gly Cys Ser Arg Glu Thr Val Gly Arg Ile Leu Lys
177
CP75 Leu Gln Gln Phe Ala Glu Leu Glu Gly Val Ser Glu Arg Thr Ala Tyr Arg Trp Thr Thr
25
CP76 Met Ala Lys Leu Ala Glu Arg Ala Gly Met Asn Val Gln Thr Leu Arg Asn Lys Leu Asn
33
P2 C Arg Gln Gln Leu Ala Asp Leu Thr Gly Val Pro Tyr Gly Thr Leu Ser His Tyr Glu Ser
28
P2 Cox Tyr Gln Glu Phe Ala Lys Leu Ile Gly Lys Ser Thr Gly Ala Val Arg Arg Met Ile Asp
22
Helix 2
Helix 3

is consistent with the conclusion of Sauer et al., (1983) that neither Arc or Mnt proteins show any significant amino sequence homology with the DNA-binding region of λ Cro repressor. Arc and Mnt proteins do share some amino acid homology between themselves (but not with the 186 cI repressor) and do have some α helical content (cited in Pabo and Sauer, 1984).

Proteins can have structural similarities even though they show no amino acid homology (Matthews et al., 1981) and the DNA binding mechanism of the Arc, Mnt and 186 cI repressors may be similar to the λ cI, λ Cro and E. coli CAP proteins. Alternatively, a different mechanism of site-specific DNA-binding may be employed in these proteins, which warrants structural and DNA-binding studies of the 186 cI repressor.

8.2.3 RecA Cleavage Site

Phage 186 was originally regarded as a member of the non-inducible group of coliphages. Woods and Egan (1974) subsequently showed that 186 could be induced from the prophage state by treatment with nalidixic acid, mitomycin C or by UV-irradiation and phage 186 shares this property with the otherwise unrelated phage λ. Induction of the 186 prophage, like that of λ, requires a recA$^+$ host to show normal levels of UV-induction or spontaneous phage production (Woods and Egan, 1974). The non-inducible phages do not require the recA$^+$ gene product to yield normal levels of spontaneous phage production (cited in Bertani and Bertani, 1971). UV-irradiation of an E. coli cell activates the RecA protein which leads to proteolytic cleavage of the repressor of recA transcription, LexA, and the maintenance repressors of the lambdoid phages λ, 434 and P22 (Horri et al., 1981; Roberts et al., 1978; Phizicky and Roberts, 1980; Little et al., 1980). The monomeric form of the λ repressor is comprised of two domains, the amino-terminal DNA-binding domain

(Sauer et al., 1979) and the carboxy-terminal domain which is involved in oligomerisation (Pabo et al., 1979). These repressors show significant carboxy-terminal amino acid sequence homology and each repressor is cleaved at an identical Ala-Gly sequence in the hinge region between the two domains (R.T. Sauer et al., 1982).

Attempts to align the amino acid sequence of the 186 cI repressor with either of the amino or carboxy-terminal domains of the lambdoid and LexA repressors failed to detect any significant amino acid homology with either of the domains and the only Ala-Gly sequence in the 186 cI repressor was located at the extreme carboxy-terminal region of the protein.

Woods and Egan (1974) isolated several tum (turbid upon mitomycin C treated plates) mutants of phage 186, which show a markedly reduced ability for UV-induction from the prophage state and are equivalent in phenotype to λ Ind^- (Tomizawa and Ogawa, 1967). Preliminary mapping experiments with four tum mutants indicate the tum mutations lie to the right of the BglII site at 79.6% and to the left of the NotI site at 89.0% (I. Lamont, personal communication) suggesting that inactivation of the 186 cI repressor may not be by direct cleavage, but by an indirect process.

8.3 Int PROTEIN

The codon usage standard of the λ cI gene detects the predicted 186 int gene reading frame (Fig. 4.1) as do the codon usage standards of the λ int and P2 int genes (data not shown). A protein of 38.9 Kd was predicted and a candidate for the Int protein was identified in the maxicell system, using plasmid clones pEC500 and pEC501 (see Section 5.2.4). Amber mutants of the 186 int gene have now been isolated (D. Dodd, personal communication) which will allow positive identification of the int gene and its protein product.

186 and P2 Int proteins differ in size by one amino acid

(336 and 337 amino acids respectively) and have very similar frequencies of basic (both 16.3%), acidic (12.5% and 11.6% respectively) and hydrophobic residues (44.1% and 45.4% respectively). Both proteins are highly basic, comparable with the basicity of the λ Int protein (186, P2 and λ Int at 16.3%, 16.3% and 16.6% basic respectively). Modulating codons are present at a high frequency (186, P2 and λ at 7.1%, 9.8% and 5.2% respectively) suggesting that all three proteins show slow translational elongation rates.

The 186 and P2 Int proteins appear to be remarkably similar judging by the predicted amino acid composition of the proteins. Figure 8.2 is a comparison of the DNA and amino acid sequences of the two proteins. DNA and amino acid homology was predominantly at the amino-terminal and carboxy-terminal regions of the proteins. The homology was increased by introducing a nine amino acid displacement and a single amino acid (Pro CCT) deletion into the P2 Int sequence. Whether the absence of homology in the central section of the two proteins is of biological significance is not known.

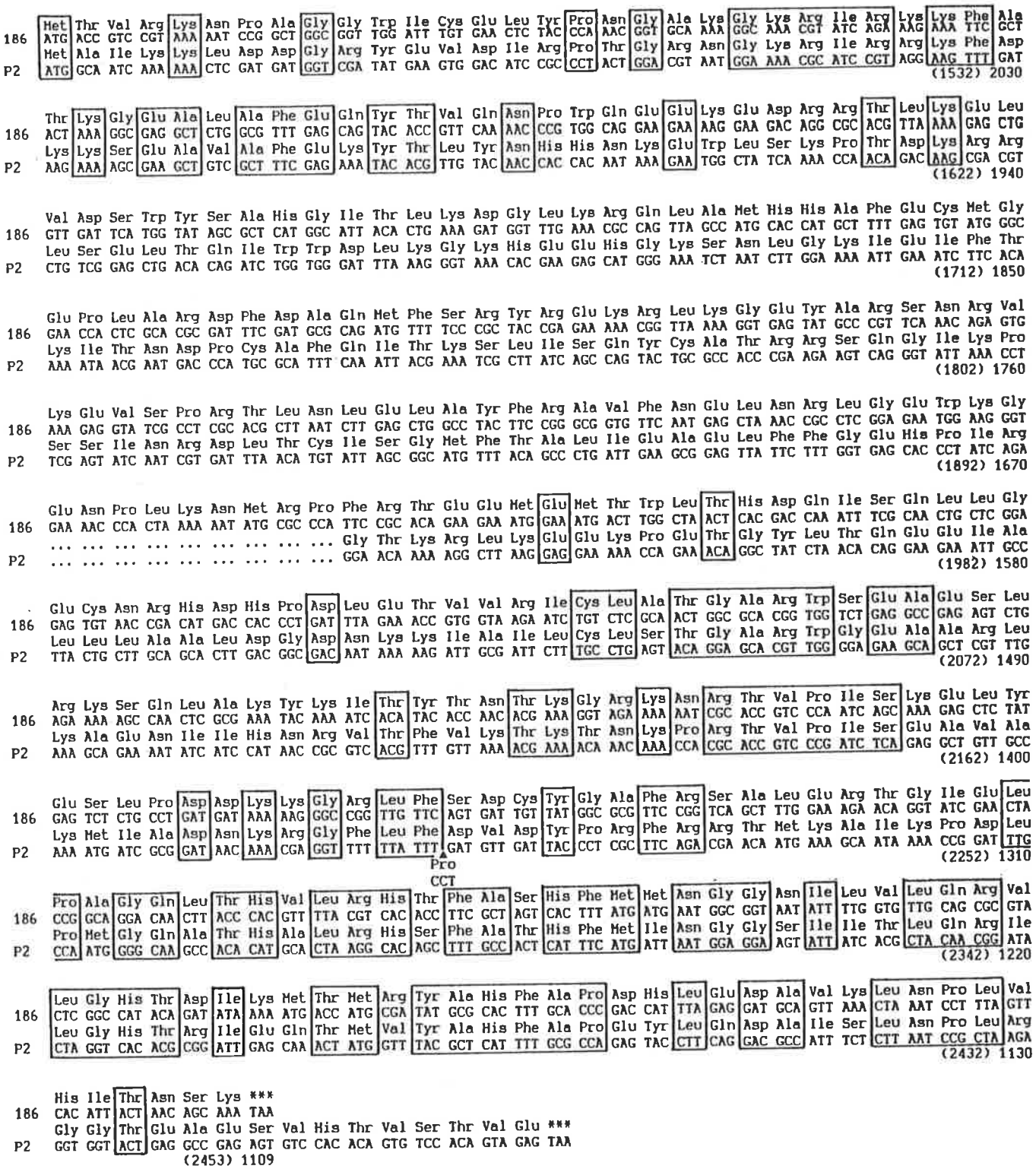
Preliminary comparisons of the Int proteins of 186, P2, λ and P22 have revealed some regions of homology, suggesting that these proteins may be related (A. Landy, personal communication).

8.4 CP69

CP69 was detected using the codon usage standard of the λ cI gene (Fig. 4.1) and was predicted to be a protein of 83 amino acids (8.4 Kd). The protein is rich in proline residues (Table 8.1), has a low polarity index (Capaldi and Vanderkooi, 1972) of 29%, and is thus considered a potential membrane-associated protein. Modulating codons are present at a high frequency (14.5%, Table 8.1), suggesting a slow translational elongation rate. A protein of 8.4 Kd was identified as a candidate for the CP69 protein in the maxicell system using plasmid clone pEC500 and pEC501 (see

Fig. 8.2. Amino acid sequence homology of the 186 and P2 int genes

Amino acid sequence homology between the predicted 186 *int* and P2 *int* genes is boxed. The DNA sequence of the P2 *int* gene was kindly provided by E. Ljungquist (personal communication). The P2 *int* sequence is displaced to optimise the amino acid homology. Numbering of the DNA sequence for the 186 *int* gene is from the *r*-strand and the numbers in parenthesis refer to the corresponding sequence position in the *l*-strand.



Section 5.2.5).

The function of CP69 is not known but must be non-essential since it is removed in the plaque-forming deletion mutant Δ1, (Section 4.2.7). CP69 was predicted to be expressed from the 186 prophage, together with the cI repressor gene and int gene, from the pL transcript (Section 6.1). Phage λ expresses two functions from the pRM transcript in the prophage state; the λ cI immunity repressor and rex. Rex is an exclusion function that makes the λ lysogen immune to infection by several bacteriophages, including rII mutants of phage T4 (Benzer, 1955; Howard, 1967). Sequence analysis of this region revealed two open reading frames (Landsmann et al., 1982; Sanger et al., 1982) which were named rexA and rexB. The proposed rexB gene has been suggested to have auxillary functions in both lysogenisation and lytic development (Landsmann et al., 1982). The rexB gene-product, like the CP69 gene-product, is predicted to be highly hydrophobic (both rexB and CP69 proteins have very low polarity indexes of 29%; See Section 8.1.2) and is likely to have an affinity for the bacterial membrane. CP69 and rexB show no DNA or amino acid sequence homology and differ in their predicted size (8.4 Kd and 16.0 Kd respectively).

8.5 CP76 AND CP75

CP76 is only partly represented on the DNA sequence of the PstI(65.5%-77.4%) fragment and the size of the protein product of this potential gene is at least 147 amino acids. The DNA sequence of the PstI-BglII(77.4%-79.6%) fragment has been determined (Richardson, H., Kalionis, B. and Egan, J.B. manuscript in preparation) and computer analysis, employing the GENE program (Section 2.23.1), detected four open reading frames with the codon usage standard of E. coli genes. One of these frames extended 22 amino acids from the PstI site at 77.4% and was contiguous with the reading frame

of the CP76 protein predicted from the DNA sequence of the PstI (65.5%-77.4%) fragment. The length of the CP76 reading frame was therefore predicted to be 169 amino acids and the gene-product of CP76 was predicted to be a protein of 18.7 Kd.

A plasmid clone (pEC402) of the XhoI-BglII(67.6%-79.6%), region which contains the Δ1 deletion (Fig. 5.2), was used in the maxicell system of Sancar et al. (1979) and directed the synthesis of several proteins, one of which was sized at 9.8 Kd and considered a potential candidate for the 9.8 Kd CP75 gene product (Section 5.2.5). The synthesis of four other proteins was also directed by the XhoI-BglII(67.6%-79.6%) DNA insert and one of these was a protein of 18.7 Kd (A. Puspurs, unpublished data). Computer prediction of the size of the CP76 protein using the MWCALC program (Staden, 1980) was 18.7 Kd. The 18.7 Kd protein observed in the maxicell system was therefore considered a candidate for the CP76 gene-product.

A noticeable feature of the CP75 and CP76 proteins was the low frequency of modulating codons (1.0% and 0.6% respectively), suggesting that both proteins would show rapid translational elongation rates. CP75 is also noticeably more basic than CP76 with 16.1% and 8.9% basic amino acid content respectively.

The function of predicted genes CP75 and CP76 is not known. Genes in analogous positions in the related phage P2, are cox, which is required for excision (Lindahl and Sunshine, 1972) and the replication gene P2 B. DNA sequencing of the region encoding these two P2 genes has been completed and both genes have been identified by the sequencing of mutants (E. Ljungquist, personal communication). The P2 DNA sequence of this region (kindly supplied by E. Ljungquist) was analysed and no significant DNA or amino acid sequence homology between CP75 and cox or between CP76 and P2 B was detected. All other possible reading frames in these regions

of the two phage were compared and again no significant homology was detected. Neither CP75 nor cox showed any significant homology with the λ excisionase gene xis.

8.6 D PROTEIN

The D gene is only partly represented on the DNA sequence of the PstI(65.5%-77.4%) fragment since two alleles of the D gene map to the left of the PstI(65.5%) site (Finnegan and Egan, 1979). DNA sequence analysis of the P2 ogr promoter-regulatory region (Section 7.4.3) revealed an open reading frame which showed extensive amino acid homology with the 186 D protein sequence. The homology was also evident at the DNA sequence level and was not unexpected, since heteroduplex analysis (Younghusband and Inman, 1974) indicated that many of the morphological genes of 186 and P2 were at least partially homologous. P2 tail gene D maps to the immediate left of the ogr gene (Bertani et al., 1982) and so the expectation is that the open reading frame of P2 showing homology to the 186 D protein, is that of the P2 D tail protein.

8.7 B PROTEIN

The B protein is of considerable interest as this protein controls phage 186 late gene transcription and, either directly or indirectly, plays a role in the turn off of early gene transcription and DNA replication (Finnegan and Egan, 1981, See Section 1.5.2). Additionally, B protein may require replicating DNA as a template in order to effect the control of late gene transcription (See Section 9.3.2). The related phages P2 and P4 encode proteins Ogr and δ respectively, which act analogously to 186 B to control late gene transcription.

The P2 Ogr protein controls P2 late gene expression and is presumed to function by activating P2 late promoters (Sunshine and Sauer, 1975; Christie and Calendar, 1983, 1985). Phage 186

B protein presumably activates these same promoters, as shown by the existence of viable P2.186 hybrid phages (Hy 9, 10, 11 - Hocking and Egan, 1982d) where P2 ogr is absent and P2 late gene expression is controlled by the 186 B gene. Satellite phage P4 δ gene-product (Souza et al., 1977) has been shown to act at the late promoters of P2 during transactivation of the helper phage P2 (Christie and Calendar, 1983, 1985) and presumably transactivates the 186 late promoters when phage 186 is used as a helper (B. Sauer et al., 1982). The expectation was that the gene-products of 186 B, P2 ogr and P4 δ are functionally related and would share some amino acid homology. DNA sequences of all three genes were available and provided the opportunity to examine both amino acid and DNA sequence homology between the three proteins.

The amino acid sequence gained by conceptual translation of the DNA sequence of P2 gene ogr (Pritchard, 1984; Pritchard and Egan, 1985; Ljungquist and Christie, manuscript in preparation; this work Section 7.4.3) was compared with that gained from the 186 B gene (Section 5.2.2., Fig. 4.5) and the P4 δ gene (Lin, 1984 and corrected by C. Halling, personal communication). The comparison is presented in Figure 8.3.

Phage 186 gene B and P2 gene ogr code for proteins each of 72 amino acids, whilst the protein of P4 gene δ contains 166 amino acids. Late gene expression of phages 186 and P2 is normally dependent on their respective replication functions (186 A and P2 A, B) (Finnegan and Egan, 1981; Lindahl, 1970; Lindqvist, 1971) and the larger size of the P4 δ gene product may reflect its ability to activate late gene expression of either helper phage, independently of the replication functions of either helper phage during transactivation. (B. Sauer et al., 1982; Six and Lindqvist, 1971; Six, 1975).

As shown in Figure 8.3, the three proteins are related at both the DNA and amino acid sequence level. Phage 186 B protein and P2 Ogr protein share considerable amino acid homology, with 32 out of 39 amino-terminal residues identical. The homology lessens for the remaining 33 carboxy-terminal amino acids with only 13 identical residues. B protein and P4 δ protein show less homology with 18 identical residues among the 39 amino-terminal residues, and only 2 in the next 33. P4 δ protein shares 19 of the 39 amino-terminal residues with P2 Ogr protein and only 3 residues in the next 33.

At the nucleotide sequence level 186 and P2 late control genes B and ogr show extensive homology, differing in only 72 of the 216 residues, and 20 of these 72 represent a change in the third base of a codon that nevertheless conserves amino acid identity. Considering the first 216 nucleotides of the P4 δ gene, 89 nucleotides are found in common with the 186 B gene, and 84 in common with the P2 gene ogr.

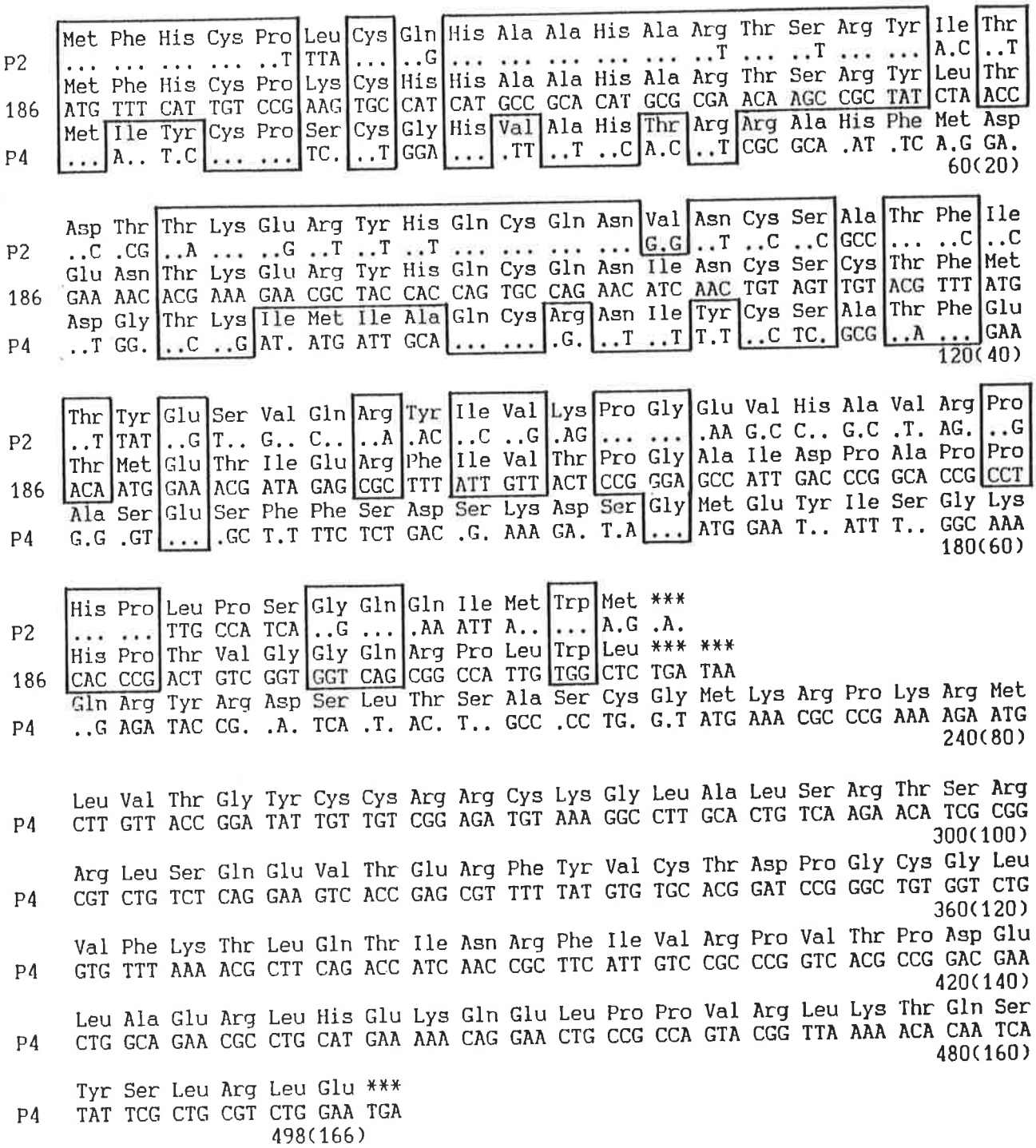
Other interesting features of the amino acid composition of these proteins include the clustering of proline residues in the carboxy-terminal sequence of 186 B protein (residues No. 52, 57, 59, 60, 62, 69) and of P2 Ogr protein (residues No. 52, 60, 62, 64), and the conservation of cysteine in the amino terminal sequence of the three proteins (residues No. 4, 7, 30, 35). Cysteine appears at a frequency of 1.8% in the amino acid composition of all coding sequences in the EMBL sequence library (Staden, 1984a), and appears at a frequency of 6.9%, 5.6% and 6.0% for the B, Ogr and δ proteins respectively, suggesting these are cysteine rich proteins and show the potential for a high degree of covalent cross-linking. P2 ogr, 186 B and P4 δ genes all show a high percentage of modulating codons (5.6%, 5.6% and 6.6% respectively) and would

Fig. 8.3. DNA and amino acid homology of the genes P2 ogr and P4 δ with the 186 B gene

Amino acid homology to the 186 B gene is indicated by the boxes and only the nucleotides which differ to the 186 B DNA sequence are presented.

For the purposes of this comparison, the numbering refers to the distance in nucleotides from the A residue of the predicted initiation codon of each protein. Numerals in parenthesis indicate the number of amino acids from the predicted initiating methionine.

The P4 δ sequence was from Lin (1984) and the P2 ogr sequence was from Ljungquist and Christie (manuscript in preparation), Pritchard and Egan (1985), and this work (Section 7.4.3).


Met Phe His Cys Pro Leu Cys Gln His Ala Ala His Ala Arg Thr Ser Arg Tyr Ile Thr
P2 ... ...T TTA ... ..G ... ...T ... ..T ... ... A.C ..T
Met Phe His Cys Pro Lys Cys His His Ala Ala His Ala Arg Thr Ser Arg Tyr Leu Thr
186 ATG TTT CAT TGT CCG AAG TGC CAT CAT GCC GCA CAT GCG CGA ACA AGC CGC TAT CTA ACC
Met Ile Tyr Cys Pro Ser Cys Gly His Val Ala His Thr Arg Arg Ala His Phe Met Asp
P4 ... A.. T.C ... ... TC. ..T GGA ... .TT ..T ..C A.C ..T CGC GCA .AT .TC A.G GA.
60(20)
Asp Thr Thr Lys Glu Arg Tyr His Gln Cys Gln Asn Val Asn Cys Ser Ala Thr Phe Ile
P2 ..C .CG ..A ... ..G ..T ..T ..T ... ... G.G ..T ..C ..C GCC ... ..C ..C
Glu Asn Thr Lys Glu Arg Tyr His Gln Cys Gln Asn Ile Asn Cys Ser Cys Thr Phe Met
186 GAA AAC ACG AAA GAA CGC TAC CAC CAG TGC CAG AAC ATC AAC TGT AGT TGT ACG TTT ATG
Asp Gly Thr Lys Ile Met Ile Ala Gln Cys Arg Asn Ile Tyr Cys Ser Ala Thr Phe Glu
P4 ..T GG. ..C ..G AT. ATG ATT GCA ... ... .G. ..T ..T T.T ..C TC. GCG ..A ... GAA
120(40)
Thr Tyr Glu Ser Val Gln Arg Tyr Ile Val Lys Pro Gly Glu Val His Ala Val Arg Pro
P2 ..T TAT ..G T.. G.. C.. ..A .AC ..C ..G .AG ... ... .AA G.C C.. G.C .T. AG. ..G
Thr Met Glu Thr Ile Glu Arg Phe Ile Val Thr Pro Gly Ala Ile Asp Pro Ala Pro Pro
186 ACA ATG GAA ACG ATA GAG CGC TTT ATT GTT ACT CCG GGA GCC ATT GAC CCG GCA CCG CCT
Ala Ser Glu Ser Phe Phe Ser Asp Ser Lys Asp Ser Gly Met Glu Tyr Ile Ser Gly Lys
P4 G.G .GT ... .GC T.T TTC TCT GAC .G. AAA GA. T.A ... ATG GAA T.. ATT T.. GGC AAA
180(60)
His Pro Leu Pro Ser Gly Gln Gln Ile Met Trp Met ***
P2 ... ... TTG CCA TCA ..G ... .AA ATT A.. ... A.G .A.
His Pro Thr Val Gly Gly Gln Arg Pro Leu Trp Leu *** ***
186 CAC CCG ACT GTC GGT GGT CAG CGG CCA TTG TGG CTC TGA TAA
Gln Arg Tyr Arg Asp Ser Leu Thr Ser Ala Ser Cys Gly Met Lys Arg Pro Lys Arg Met
P4 ..G AGA TAC CG. .A. TCA .T. AC. T.. GCC .CC TG. G.T ATG AAA CGC CCG AAA AGA ATG
240(80)
Leu Val Thr Gly Tyr Cys Cys Arg Arg Cys Lys Gly Leu Ala Leu Ser Arg Thr Ser Arg
P4 CTT GTT ACC GGA TAT TGT TGT CGG AGA TGT AAA GGC CTT GCA CTG TCA AGA ACA TCG CGG
300(100)
Arg Leu Ser Gln Glu Val Thr Glu Arg Phe Tyr Val Cys Thr Asp Pro Gly Cys Gly Leu
P4 CGT CTG TCT CAG GAA GTC ACC GAG CGT TTT TAT GTG TGC ACG GAT CCG GGC TGT GGT CTG
360(120)
Val Phe Lys Thr Leu Gln Thr Ile Asn Arg Phe Ile Val Arg Pro Val Thr Pro Asp Glu
P4 GTG TTT AAA ACG CTT CAG ACC ATC AAC CGC TTC ATT GTC CGC CCG GTC ACG CCG GAC GAA
420(140)
Leu Ala Glu Arg Leu His Glu Lys Gln Glu Leu Pro Pro Val Arg Leu Lys Thr Gln Ser
P4 CTG GCA GAA CGC CTG CAT GAA AAA CAG GAA CTG CCG CCA GTA CGG TTA AAA ACA CAA TCA
480(160)
Tyr Ser Leu Arg Leu Glu ***
P4 TAT TCG CTG CGT CTG GAA TGA
498(166)

be expected to show slow translational elongation rates.

P4 δ protein is noticeably more basic (16.3%) than the B and Ogr proteins (both 9.7%) and unlike the scattered distribution of the basic amino acids in B and Ogr, the basic amino acids of P4 δ protein tend to be clustered at the carboxy-terminus of the protein, with 8 basic residues in the first 72 amino acids and 19 basic residues (predominantly arginine) in the last 94 amino acids. The high basic amino acid content and clustering of these residues indicates a possible interaction of this region of the protein with phosphate backbone of DNA (Section 8.1.4).

The three control proteins, B, Ogr and δ, score below the threshold of detection for the α2-α3 DNA-binding helix motif (Table 8.2).

SECTION 9

GENERAL DISCUSSION : SYSTEMS OF CONTROL

9.1 CONTROL OF LYTIC AND LYSOGENIC FUNCTIONS

9.1.1 Control of Lytic and Lysogenic Functions in Phages λ and 186

As described in detail previously (Section 1.2), the control of lysis and lysogeny in phage λ can be viewed as consisting of two genetic switches. The first switch determines whether lytic or lysogenic functions are expressed and the second switch determines whether the phage DNA is integrated into the bacterial chromosome or remains free in the cytoplasm. These two switches must be co-ordinately controlled to ensure that during lytic expression the phage DNA remains free in the cytoplasm, whilst during lysogenic expression the phage DNA is integrated into the host chromosome.

Phage 186 is unrelated to λ, but it is also a temperate phage with the ability to carry out the lytic or lysogenic pathways of development. The arrangement of genes differs markedly in many important respects to that of λ (Section 1.3) and indicates a different control strategy to that of phage λ. The potential therefore exists for novel control mechanisms, but also of importance would be the identification of common regulatory features between these unrelated phages which may define the minimal requirements for a temperate phage to carry out the lysis or lysogeny decision.

9.1.2 Organisation of Lytic and Lysogenic Functions in Phage λ and 186

The organisation of the lytic and lysogenic functions of phage λ was compared in detail with that of phage 186 in Section 1.3. Control of lytic and lysogenic functions in λ occurs primarily at the oR operator and ultimately reflects the outcome of the competition between two repressors cI and Cro for binding at the oR operator region. Binding of the Cro repressor to oR blocks transcription of the cI gene allowing lytic development to proceed; conversely

binding of cI repressor blocks transcription of the cro gene allowing lysogenic development. Johnson et al. (1981) have described the following essential features of this reciprocal repression:-

1) An operator site oR between the two repressor genes, where the two repressors bind.

2) This operator site is tripartite (i.e. three adjacent repressor binding sites) and the cI and Cro repressors have opposite affinities for the three binding sites. The promoter of cI transcription (pRM) and that of the cro gene (pR), overlap opposite ends of the operator.

3) The co-operative nature of the cI repressor binding at the oR operator region, which stimulates pRM transcription and thus leads to a rapid increase in repressor concentration.

Genetic studies of phage 186 (Bradley et al., 1975; Hocking and Egan, 1982a) have identified a cI gene (analogous to λ cI) which is required for the maintenance of lysogeny, but a cro-like function in phage 186 has not yet been identified. Mutation to virulence (determined by the ability to form plaques on a lysogen) occurs relatively frequently (10^{-8}-10^{-10}) and therefore a single operator site was expected in phage 186.

The overall organisation of lytic and lysogenic functions of phage 186 appears to be considerably less complex than that of phage λ. Phage 186 contains two promoters which are expected to be essential in the lysis/lysogeny decision, one of which is required for expression of the lysogenic functions (pL) and the other for expression of lytic functions (pR). The pL promoter was predicted to be a weaker promoter than pR (Section 4.2.3). Promoter pL was shown to be active in vivo (Section 6.2.3) whilst promoter pR was active both in vivo (Section 6.2.2) and in vitro (Pritchard, 1984; Pritchard and Egan, 1985). Transcription from these two promoters converges and overlaps in the operator region, unlike transcription

from the lysogenic (pRM) and lytic (pR) promoters of phage λ. This suggests that 186 has the potential for different mechanisms to sustain the lysis/lysogeny states (establishment, maintenance, induction, lytic infection) to those operating in phage λ. Further understanding of how the lysis/lysogeny states might be sustained in phage 186 required the identification of the operator region.

9.1.3 The 186 Operator Region and the Arrangement of the Lytic and Lysogenic Promoters

Finnegan and Egan (1981) showed that early lytic transcription was under cI repressor control and located the operator to the immediate right of the cI gene, at about 75% from the left hand end of the 186 genome. DNA sequence analysis of the 130 bp region between the initiation codon of the cI gene and CP75 (the first predicted gene on the early lytic transcript) revealed the presence of three direct repeats (OD1, OD2, OD3) and three inverted repeats (OI1, OI2, OI3), giving two candidates for a tripartite operator in phage 186 (Section 7.1).

DNA sequence analysis of 186 virulent mutants revealed multiple base-pair changes in the presumptive operator region and each virulent mutant was found to have at least one base-pair change in each of two distinct areas. The first area was within the inverted repeat OI1 and, since several virulent mutants did not have any base-pair changes within the three direct repeats (OD1, OD2, OD3), it was concluded that inverted repeat OI1 (and by implication OI2 and OI3) represents a cI repressor protein binding site. Notice should be taken of the weak inverted repeat nature of OI2 and OI3 which is essentially restricted to the central palindrome (5'-CAATTG-3') found in OI1. OI1 overlaps both -10 and -35 regions of the pR early lytic promoter and if OI1 is a repressor binding site then the pR promoter is under direct cI repressor control. Although direct

repeats OD1, OD2 and OD3, are not considered to be candidates for cI repressor protein binding sites, the possibility that these repeats represent binding sites for some as yet unidentified protein involved in the control of the lysis/lysogeny decision cannot be excluded.

The second area of base-pair changes in the operator region was in the DNA sequence from the predicted cI ribosome binding site to the initiation codon of the cI gene. These base-pair changes could potentially result in reduced expression of the cI repressor gene by interfering with translation, but if this were an essential feature for the virulence phenotype then mutations in the cI coding region would equally well have been expected. All nine virulent mutants sequenced over the operator region contained at least one base-pair change in this second area (Section 7.2). Whether this area forms an essential part of the cI repressor binding site or represents a second binding site for an as yet unidentified protein involved in the control of lytic gene expression is not known.

The most striking feature of the operator region is the arrangement of the two opposing promoters pR and pL, which are expected to be essential for the expression of lytic and lysogenic functions respectively. The consequences of such an arrangement, with regard to the means by which the lytic and lysogenic states could be sustained in phage 186, are considered in the following discussion.

9.1.4 The Lytic State

Ward and Murray (1979) provided evidence that convergent transcription between the E. coli trp and λ pL promoters resulted in the mutual impairment of gene expression directed by both promoters. They proposed that RNA polymerase molecules moving in opposing directions can collide, resulting in transcription termination

and they also showed that transcriptional interference from a strong promoter can completely block gene expression from a weaker promoter.

The possibility therefore existed that lytic development of phage 186 could be sustained by transcription from the strong pR lytic promoter blocking cI gene expression directed by the weaker opposing pL promoter (See also Section 4.2.3).

Towards testing this proposal, the 258 bp TaqI fragment (sequence co-ordinates 2700-2957) containing promoters pL and pR, was cloned into the promoter expression vector pKO1 (McKenney et al., 1981) and galK expression, under control of the pL promoter, was measured (I. Dodd, personal communication). GalK expression of reasonably high levels was detected from the pL promoter even though the stronger pR promoter was being actively transcribed from the opposite direction (I. Dodd, personal communication). Although the possibility of some interference of pR transcription with transcription from the pL promoter has not been excluded, the above experiment suggests that pR transcription in the lytic state is not by itself sufficient to completely block pL transcription.

Presumably, some means of preventing cI expression during lytic development must exist to prevent the possibility of re-establishment of repression after committment to the lytic cycle of development. Phage λ prevents cI expression by the action of the Cro repressor binding to operator oR3 and blocking transcription from the pRM promoter (and hence cI transcription). The possibility existed that phage 186 might contain a cro-like function. However, Finnegan and Egan (1981) showed that the int-cI region was transcriptionally active throughout lytic development following the heat induction of a 186cItsp prophage. Although by no means conclusive, this data is not compatible with the existence of a phage-encoded cro-like function which acts to prevent pL transcription

and hence cI gene expression. Prevention of cI expression could also possibly occur by translational or post-translational control. An example of post-translational control of repressor activity is the antirepressor protein (Ant) of temperate phage P22, which acts as an antagonist of the primary c2 phage repressor (reviewed by Susskind and Youderian, 1983).

9.1.5 The Lysogenic State

The basic requirements for the maintenance of lysogeny are that the repressor protein blocks transcription of the early lytic promoter(s), whilst allowing its own transcription. Maintenance of lysogeny in temperate phage λ involves both positive and negative control of transcription of the cI maintenance repressor gene from the pRM promoter (See Section 1.2). The arrangement of opposing pL (lysogenic) and pR (lytic) promoters in the operator region of phage 186 has important consequences which relate to the manner in which the prophage state is maintained.

An RNA polymerase molecule initiating transcription of the cI gene from the pL promoter must be capable of transcribing past the cI repressor protein, which is blocking pR transcription by binding to the DNA at the operator site, to allow the continued transcription of the cI gene and thereby maintain the lysogenic state. Evidence that this situation does occur comes from experiments carried out by I. Dodd (personal communication) using the promoter expression vector pKO1 (McKenney et al., 1981). A plasmid clone of the 1277 bp HaeIII fragment (sequence co-ordinates 1723 - 2999) containing the cI gene (but only part of int and CP75) and the operator region, was found to confer immunity to superinfection at the permissive temperature (I. Dodd, personal communication). This suggested that synthesis of the cI repressor occurred from pL. Promoter pR-controlled expression of galK was not detected

in this clone (I. Dodd, personal communication), indicating that transcription from pR was being repressed by cI. Repressor protein blocking transcription from the pR promoter therefore does not appear to prevent transcription from the opposing pL promoter and so allows the maintenance of the prophage state.

9.1.6 The P2 Operator Region and the Arrangement of the Lytic and Lysogenic Promoters

Control of the lytic and lysogenic states of phage 186 appears to occur primarily in the operator region and it was therefore of interest to compare this region with the operator region of the related phage P2.

Phage P2, like its close relative 186, is believed to contain a single operator region where the P2-specific repressor C binds to regulate the expression of the early lytic genes cox, B and A (reviewed in Bertani and Bertani, 1971). The DNA sequence of the P2 C maintenance repressor gene and the operator region of phage P2 were determined by Ljungquist et al. (1984).

DNA sequence analysis of three virulent deletion mutants led to the prediction that two 8 bp direct repeat sequences, O1 and O2 constitute at least part of the P2 repressor-binding sites (Ljungquist et al., 1984). The P2 operator region also carries three inverted repeat sequences which share homology and for the purposes of this discussion were named O1', O2' and O3'. O1' and O2' were noted by Ljungquist et al. (1984) and O3' was detected by I. Dodd (personal communication). These inverted repeat sequences are presented in Table 9.1. The locations of the inverted repeats (O1', O2') and the direct repeats (O1, O2) are consistent with the position of the virulent deletion mutants (Ljungquist et al., 1984) and therefore both the inverted and direct repeats must be considered candidates for the P2 operator sites.

Table 9.1 Inverted repeat DNA sequences in the operator regions of phages 186 and P2

		Spacing	Sequence[a] Position
(a) 186 operator			
OI1	TACTATCTCTCAA.TTGGGAGATATAT	32[b]	2724
OI2	TAAACCCACGCAA.TTGATGGCAAGTG	35	2756
OI3	AGAGTCAAATCAA.TTGCAAACTTTGG		2791
(b) P2 operator			
O1′	TGGTGTTTAGA.TCTCAATAGTA	22[c]	-10[d]
O2′	TTTAGTTTAGA.TGTAGATTGTT	22	-32
O3′	TAGTGCTTGGA.TGTGGGCACTA		-54

Notes to Table 9.1

a. Sequence position of the base referred to in notes b and c.
b. Number of bases between the C residue of the central palindrome CAATTG, to the C of the next palindrome.
c. Number of bases between the A residues nearest the centre of each inverted repeat.
d. Sequence position of the A residue at the centre of the inverted repeat found in the L-strand of the operator region of P2 (Ljungquist et al., 1984).

Inverted repeat sequences constitute the operator sites of phages λ, P22, φ21 (Pirotta, 1975; Humayan et al., 1977; Maniatis et al., 1975) and the DNA-binding sequences recognised by many DNA-binding regulatory proteins (Gicquel-Sanzey and Cossart, 1982). An inverted repeat sequence (OI1) was also predicted to constitute, at least in part, the 186 operator site (discussed in Section 9.1.3) and therefore the inverted repeats (O1', O2', O3') found in the P2 operator region are favoured as the P2 repressor-binding sites.

Ljungquist et al. (1984) noted a possible rightward promoter for P2 early lytic transcription (analogous to the 186 pR promoter) spanning the proposed O1 operator site, but the DNA sequence of this promoter was not given. The promoter weight matrix was used in the SCAN program (Section 2.23.2) and a promoter sequence which spanned the proposed operator site O1 (and O1') was detected (Table 9.2, Fig. 9.1). This promoter was predicted to be under direct C repressor protein control and was named P2pR for the purposes of this discussion. A promoter sequence for the P2 C repressor gene was predicted by Ljungquist et al. (1984) approximately 60 bp from the predicted initiation codon of the C gene (analogous to the 186 pL promoter) and named P2pL for the purposes of this discussion.

Promoter P2pL was suggested to be strong (Ljungquist et al., 1984), but the spacing between the predicted -10 and -35 hexamers of the P2pL promoter was only 14 bp, a spacing not found in any of the 112 promoters compiled by Hawley and McClure (1983). The promoter weight matrix, when used in the SCAN program, resulted in the detection of a more likely -35 hexamer CTCATG for the P2pL promoter, with the optimal 17 bp spacing to the -10 hexamer (TATAAT) predicted by Ljungquist et al. (1984). Promoter strengths, predicted

Table 9.2 Promoter sequences in the operator regions of phages 186 and P2

	Sequence Position[b]	Promoter Score[c]	-35	-10
Consensus[a]			TTGACa	TAtAaT
(a) 186 promoters				
pR	2740	1.2	TTTACT---17---TATATT	
pL	2817	0.7	TTGCGA---17---CATGAT	
(b) P2 promoters				
P2pR	-13[d]	1.8	TTGACA---17---TAGTAT	
P2pL	-72	-0.1	CTCATG---17---TATAAT	

Notes to Table 9.2

a. The consensus sequence for the -10 and -35 regions was from the compilation of known promoters of E. coli RNA polymerase (Hawley and McClure,1983). Lower case letters indicate nucleotides that are less highly conserved.

b. The sequence position corresponding to the rightmost conserved T residue in the -10 region (Rosenberg and Court, 1979) is given.

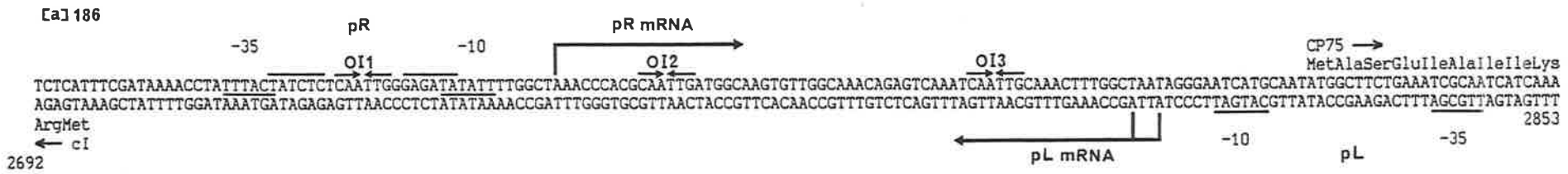
c. Promoter score was determined by the SCAN program (See Section 2.23.2). Promoter score for selected promoters from the lambda DNA sequence (Sanger et al., 1982) was 2.6 for pR, 1.3 for pL and -0.7 for pRM.

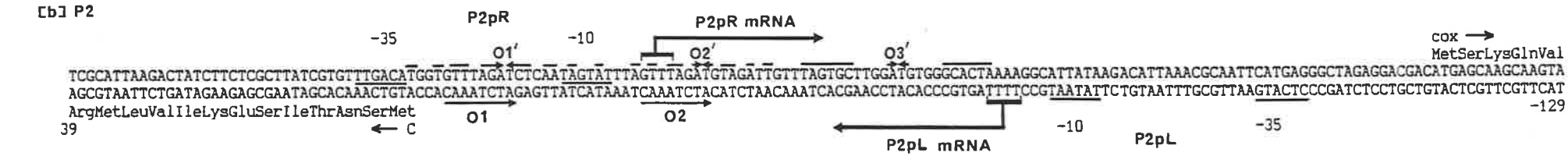
d. The P2 sequence positions refer to the DNA sequence presented by Ljungquist et al. (1984).

Fig. 9.1. Arrangement of the lytic and lysogenic promoters in the operator regions of 186 and P2

(a) The operator region of phage 186. The positions of the predicted pL (lysogenic) and pR (lytic) promoters are indicated. Inverted repeats (OI1, OI2, OI3) found in the operator region are indicated by the head to head arrows. The predicted start points of the pR and pL RNA transcripts are shown (See Section 6.2.2 and 6.2.3 respectively).

(b) The operator region of phage P2. The positions of the predicted P2pL (lysogenic) and P2pR (lytic) promoters are indicated (See Section 9.1.6). Direct repeats O1, O2 predicted to constitute at least part of the repressor-binding sites (Ljungquist et al., 1984) are shown by the arrows. The inverted repeats O1', O2' and O3' (See Section 9.1.6) are shown by the head to head arrows. The P2pL and P2pR transcripts are predicted to initiate 4 to 7 bp after the -10 hexamers (Rosenberg and Court, 1979) of the P2pL and P2pR promoters respectively.


[a] 186
pR
pR mRNA
-35
-10
OI1
OI2
OI3
CP75
MetAlaSerGluIleAlaIleIleLys
TCTCATTTCGATAAAACCTATTTACTATCTCTCAATTGGGAGATATATTTTGGCTAAACCCACGCAATTGATGGCAAGTGTTGGCAAACAGAGTCAAATCAATTGCAAACTTTGGCTAATAGGGAATCATGCAATATGGCTTCTGAAATCGCAATCATCAAA
AGAGTAAAGCTATTTTGGATAAATGATAGAGAGTTAACCCTCTATATAAAACCGATTTGGGTGCGTTAACTACCGTTCACAACCGTTTGTCTCAGTTTAGTTAACGTTTGAAACCGATTATCCCTTAGTACGTTATACCGAAGACTTTAGCGTTAGTAGTTT
ArgMet
cI
2692
2853
pL mRNA
-10
-35
pL


[b] P2
P2pR
P2pR mRNA
-35
-10
O1′
O2′
O3′
cox
MetSerLysGlnVal
TCGCATTAAGACTATCTTCTCGCTTATCGTGTTTGACATGGTGTTTAGATCTCAATAGTATTTAGTTTAGATGTAGATTGTTTAGTGCTTGGATGTGGGCACTAAAAGGCATTATAAGACATTAAACGCAATTCATGAGGGCTAGAGGACGACATGAGCAAGCAAGTA
AGCGTAATTCTGATAGAAGAGCGAATAGCACAAACTGTACCACAAATCTAGAGTTATCATAAATCAAATCTACATCTAACAAATCACGAACCTACACCCGTGATTTTCCGTAATATTCTGTAATTTGCGTTAAGTACTCCCGATCTCCTGCTGTACTCGTTCGTTCAT
ArgMetLeuValIleLysGluSerIleThrAsnSerMet
C
39
-129
O1
O2
P2pL mRNA
-10
-35
P2pL

by the SCAN program (Section 2.23.2), for the pL and pR promoters of phages 186 and P2 are presented in Table 9.2. In both phages the promoter for repressor gene transcription (pL) was predicted to be weaker than the promoter of early lytic gene transcription pR.

Finally, and most importantly, the overall arrangement of the P2 promoters for repressor gene transcription and early lytic transcription, relative to the predicted operator site(s), is very similar to that of phage 186. Transcription from the pR and pL promoters in both phages is predicted to converge and then overlap in the operator region (Fig. 9.1). Confirmation of this arrangement of promoters in the operator region of phage P2 awaits *in vivo* and *in vitro* transcription studies similar to those used to identify the 186 pL and pR promoters (Section 6).

A similar arrangement of opposing promoters for lytic and lysogenic functions in the operator region has also been reported in the unrelated temperate phage Mu (Goosen *et al.*, 1984) and suggests that this may be a common control feature among the non-lambdoid temperate phages.

9.1.7 Future Work

Transcription from the pL promoter is not completely blocked by transcription from the opposing pR promoter (Section 9.1.4), but the degree to which interference of transcription does occur should be investigated. Promoters pL and pR can be cloned individually into the promoter expression vector pKO1 (McKenney *et al.*, 1981) where their strength can be quantitated by assaying galactokinase activity. A comparison of their individual strengths with that found for the two promoters when they are present on the same DNA fragment and opposing each other's transcription, should provide a measure of transcriptional interference.

The operator site(s) requires further definition and whilst the DNA sequence analysis of virulent mutants was the simplest approach to defining the operator site(s), the pattern of base-pair changes in these mutants was unexpectedly complex. A second approach is to bind purified cI repressor protein to the operator region and carry out DNA protection experiments (so-called footprinting analysis) as described by Schmitz and Galas (1979). At increasing repressor concentrations the presence of binding sites with a lower affinity for the repressor should be revealed. These studies obviously require the isolation of the cI repressor, which can be achieved by cloning the gene into a high expression vector such as that of Remaut et al. (1983) and using standard purification procedures similar to those used for the λ cI repressor (Johnson et al., 1980).

Finally, the availability of the DNA sequence allows the construction of radioactive DNA probes to further characterise mRNA transcripts from the early control region and to assess the appearance of these transcripts, at different times after phage infection or prophage induction, and in various mutants.

9.2 CONTROL OF INTEGRATION AND EXCISION

9.2.1 Integration and Excision in Phage λ and 186

The Int protein of phage λ catalyses the site-specific recombination reaction required for integration of the phage DNA into the host chromosome and for excision from the prophage state. Int protein and the absence of Xis (excisionase) is needed for integration, whilst both Int and Xis are required for prophage excision. Phage 186 is also a temperate phage and can integrate and excise. The int gene is known to be required for integration since the ins3 (Int^-) insertion mutant is unable to form a lysogen without the aid of a helper (Int^+) phage. An excisionase gene (xis)

has not yet been identified but is presumed to be needed for 186 prophage excision.

Control of the integration and excision process in phage λ involves the utilisation of a variety of regulatory mechanisms (See Section 1.2) so that Int is absent during the lytic cycle, present without Xis during integration, and present with Xis during prophage excision. Much of the complexity in the control of λ *int* expression results from the fact that *int*, which is essentially a lysogenic function, is encoded on the λ *p*L lytic transcript. Control of the integration and excision processes in phage 186 has the potential for being considerably less complex than that found in λ. Phage 186 has essentially two transcription units in the early control region, the *p*L transcript for the lysogenic genes (*c*I, *int*) and the *p*R transcript for the early lytic genes (Section 6). Prophage integration can be simply regulated if the *int* and *c*I genes are expressed from the *p*L lysogenic promoter, whereas prophage excision during lytic development could be readily controlled if the predicted 186 *xis* gene was present on an early lytic transcript (See later discussion 9.2.4).

9.2.2 Int Expression in Temperate Phages λ and 186

Significant differences between *int* expression in phages 186 and λ can be identified even at this preliminary level of analysis. The *c*I and *int* genes of λ are expressed from separate operons whereas the analogous genes of 186 appear to be expressed from the same operon. Furthermore, the most important difference between the expression of the λ *int* and 186 *int* genes is that the 186 *int* gene is predicted to be expressed in a lysogen (from the *p*L transcript), whereas λ *int* expression in the λ lysogen is repressed.

An interesting feature of the 186 *int* gene is that the initation

codon for the int gene overlaps the termination codon of the cI gene (Fig. 4.5). This type of overlap could be advantageous in enhancing int gene translation. Translation initiation and termination codons which overlap have been studied in several bacterial operons (Oppenheim and Yanofsky, 1980; Schumperli et al., 1982; Aksoy et al., 1984) and evidence was provided that complete translation of the preceding gene was necessary for optimal translation of the distal gene. This phenomenon was called translational coupling. Genes expressed from polycistronic mRNAs in phage λ also show overlapping translation initiation and termination codons (Sanger et al., 1982).

The possibility therefore exists that not only will the 186 int gene be expressed in the lysogen, but its optimal translation is also predicted to be dependent on complete translation of the cI gene through translational coupling.

9.2.3 Comparison of Int Expression in Related Phages 186 and P2

Phage P2 is closely related to 186 and the two phage show many similarities in their gene organisation. The int gene of each phage is located next to the maintenance repressor gene and the availability of the DNA sequence of the int-C region of phage P2 (E. Ljungquist, personal communication) allowed a comparison of the two regions to be made. The Int proteins of the two phages showed amino acid homology in the amino-terminal and carboxy-terminal regions (Section 8.3), but in contrast the DNA sequence of the int-repressor regions of the two phages show the potential for quite different control of int gene expression as detailed in the following discussion.

Computer analysis of the P2 int-C region (Ljungquist et al., 1984; E. Ljungquist, personal communication) revealed that the distance between the termination codon of the P2 C gene and the initiation codon of the P2 int gene was 115 bp (data not shown).

Therefore, unlike the 186 int-repressor genes, the int-repressor genes of phage P2 do not appear to show translational coupling. Additionally, the DNA sequence between the P2 int-repressor genes contains a presumptive rho-dependent transcription terminator (Ljungquist et al., 1984), suggesting the possibility that, unlike the 186 int-repressor genes which were encoded on the 2.2 kb pL transcript, the P2 int gene may not be cotranscribed with the repressor gene (C). Evidence that the P2 int gene is expressed from its own promoter is described below.

P2 DNA integrated into the host chromosome forms the classic example of a non-excisable prophage state (L.E. Bertani, 1968). The inability to excise is believed to reflect the poor expression of the int gene in the prophage state (Bertani, 1970), which is known to be required for spontaneous phage production from the lysogen as well as for prophage integration (Choe, 1969). Bertani (1970) proposed the "split-operon" model to account for the poor expression of int in the prophage state which involved the splitting of the int gene from its promoter upon insertion of the phage DNA into the host chromosome. This model demanded the P2 int promoter be located to the left of the att site and that transcription of the P2 int gene was rightwards with respect to the genetic map. This model has been invalidated by determining that the int gene is in fact transcribed leftwards (Ljungquist and Bertani, 1983; E. Ljungquist, personal communication), as described in detail in Section 1.4.2.

Ljungquist and Bertani (1983) cloned the PstI(62.6%-76.7%) fragment of P2 wild type DNA into pBR322 and demonstrated functional int^+ activity of the insert when cloned in either orientation. In these studies a pBR322 clone of the BglII(51.2%-75.4%) fragment did not give functional int^+ activity. These results suggested that P2 int was transcribed from a promoter in the region 75.4%-76.7%

of the P2 region. This region spans the 115 bp P2 int-C intergenic region but computer analysis of this region using the promoter weight matrix (Section 2.23.2) did not detect a promoter sequence (DNA sequence supplied by E. Ljungquist).

DNA sequence analysis of the 80 bp region between the termination codon of the P2 int gene and the att region (DNA sequence supplied by E. Ljungquist) revealed no potential stem-loop terminator structure which could act as a terminator for int transcription. This represents a potentially important difference in expression of the int gene in these two phages since termination of transcription of the 186 int gene (and that of cI and CP69) was predicted to occur at the potential rho-independent transcription terminator tL (Section 6.1).

Another potentially important difference between 186 and P2 int expression is that P2 has no analogous function to CP69. Plasmid clone pEC501 containing the KpnI(69.1%) to SalI(73.1%) region of phage 186 directed the synthesis of a 38.9 Kd protein which was consistent in size with that predicted for the Int protein (Fig. 5.2, Section 5.2.4). The presence of this plasmid clone conferred detrimental effects on cell viability, however these effects were suppressed in the XhoI(67.6%)-SalI(73.1%) plasmid clone pEC500 (A. Puspurs, personal communication). This clone directs the synthesis of the 38.9 Kd protein and an 8.4 Kd protein consistent in size with the predicted CP69 protein. The possibility therefore exists that CP69 might be important in controlling 186 int expression.

At this stage it is not possible to distinguish whether any effect on 186 int gene expression might result from the action of CP69 protein or reflect the presence of direct repeats R1 to R6 found in the DNA sequence spanning CP69. Closely spaced clusters of direct repeats have been found near the origins of DNA replication

of phages λ, φ80 and φ82 (Tsurimoto and Matsubara, 1981; Grosschedl and Hobom, 1979; Hobom *et al.*, 1979; Moore *et al.*, 1979) and of several plasmids (reviewed by Scott, 1984). These direct repeats have been proposed to be binding sites for replication proteins and this has been confirmed to be the case in λ (Tsurimoto and Matsubara, 1981). Repeats R1 to R6 are not expected to play any similar role in 186 replication (the origin of replication *ori* is located 7.5 kb from these repeats at 93%, Section 1.3.1(b)) and the repeats are not important for the viability of the phage since they are removed in the deletion phage Δ1 (Section 8.4). The possibility does however exist that these repeats in the early control region of phage 186, represent a cluster of binding sites for some as yet unidentified protein. The close proximity of repeats R1 and R6 (located at 68.4%-69.1%) to the *att* site (approximately 68%) and the *int* gene suggests a possible involvement in the integration or excision process.

This preliminary comparison of *int* expression in phages 186 and P2 clearly indicates the potential for quite different control of *int* gene expression in these two closely related phages.

9.2.4 Prediction of a 186 Excisionase Gene and its Role in the Control of Integration and Excision

*Ins*3(*Int*$^-$) lysogens can be constructed only with the aid of a helper (*Int*$^+$) phage and these lysogens do not yield phage upon prophage induction (Bradley *et al.*, 1975). This suggests a requirement of the functional *int* gene product for the excision process as well as for integration. The involvement of the *int* gene in both these processes demands the existence of a phage-encoded function, *xis*, which presumably acts analogously to λ *xis* and alters the activity of the 186 *int* gene product to allow the excision reaction to occur. Genetic studies have not been pursued to identify

the predicted 186 xis gene but its possible location was suggested by the identification of the excision gene cox in the closely related phage P2.

Although the P2 prophage is considered non-excisable (see Section 1.4.2), spontaneous release of P2 phage from P2 lysogens does occur at a very low frequency, and both P2 genes int and cox are required for this reaction (Lindahl and Sunshine, 1972). P2 cox therefore shows the properties predicted of the 186 xis gene, in that it acts in combination with the P2 int gene to allow excision. The cox gene is the first gene on the P2 lytic operon and the sequence of this gene has been completed (E. Ljungquist, personal communication). CP75 maps in the anologous position to cox and the possibility existed that amino acid sequence homology may be shared between the two proteins. However, computer analysis failed to detect significant DNA or amino acid sequence homology and neither protein shows homology to the λ excisionase protein, Xis (Section 8.5). The lack of homology between the P2 excision gene cox and CP75 does not preclude the possibility that CP75 is the 186 xis gene and further studies are required to verify the existence of the 186 xis gene (Section 9.2.5).

Although the predicted 186 xis gene has not been identified, distinct advantages in the control of integration and excision can be anticipated if the predicted xis gene was expressed from an early lytic transcript. During prophage induction, expression of the xis gene from an early lytic transcript would ensure excision from the bacterial chromosome before lytic development and would also prevent any possibility of integration during lytic infection.

Assuming that Int alone is required for integration and both Int and Xis are required for excision, then the control of the integration and excision processes could simply rely on the differential expression of xis from an early lytic transcript.

The int gene was predicted to be expressed from the 2.2 kb pL lysogenic transcript during the maintenance of lysogeny (Section 6.1). Inactivation of the cI repressor during prophage induction would allow Xis to be produced from a lytic transcript, this would then act together with the Int protein synthesised during the lysogenic state and allow excision of the prophage from the bacterial chromosome. Expression of int during the maintenance of lysogeny could therefore be important in ensuring that prophage induction is rapid, by providing a pool of Int protein to act with Xis. During lytic development cI repression of the pR promoter is not established and the synthesis of Xis from an early lytic transcript would prevent the possibility of integration. Conversely, during the establishment of lysogeny, transcription from the pR promoter is blocked by the cI repressor and the synthesis of Xis is prevented.

Continued synthesis of Xis after prophage induction and during lytic development may indeed be essential to prevent reintegration since Finnegan and Egan (1981) showed that transcription from the cI-int region (and presumably Int protein synthesis) continues following prophage induction.

In conclusion, the regulation of integration and excision in phage 186 could potentially be less complex than that found in λ and rely primarily on the presence or absence of Xis.

9.2.5 Future Work

The primary objective for future work on the control of the integration and excision processes is to identify the predicted 186 excisionase gene. CP75 is a possible candidate for this gene since it maps in the analogous position to the P2 excision gene cox. Site-directed in vitro mutagenesis (Zoller and Smith, 1983) could be used to create CP75 mutants which can then be tested for

their ability to excise. The role of CP69 and the direct repeats R1 to R6 in the integration and excision processes is not known. Site-directed mutagenesis at the KpnI(69.1%) site, located within CP69 and R1, would determine if either play any role in integration and excision.

Northern analysis using a cI-specific probe was used to identify the 2.2 kb pL transcript which is sufficiently large enough to encode the int gene (Section 6.2.3). The use of int-specific probes should also detect the 2.2 kb pL transcript and reveal whether the int gene is expressed from any other transcript(s) in the lysogen. These probes can be used to assess the appearance of these transcripts after phage infection or prophage induction and in various mutants.

Finally, the arrangement of the 186 and P2 int genes shows the potential for quite different control (Section 9.2.3). The availability of the DNA sequence of the P2 int-repressor region will allow the construction of P2 int-specific probes to characterise P2 int gene expression by Northern analysis.

9.3 THE LATE CONTROL GENE B

9.3.1 DNA Sequence Analysis and Identification of the B Transcript(s)

DNA sequence analysis of the PstI-XhoI(65.5%-67.6%) region revealed two potential protein-coding sequences which were subsequently confirmed as the reading frames for the tail gene D (only part of the D reading frame was represented on the DNA sequence) and the late control gene B (Section 4.2.1, 5.2). Further computer analysis revealed the B gene was bounded by a transcription promoter pB and a terminator tB (Section 4.2.3, 4.2.4). A rightward transcript of 290 b was predicted for the B gene and the expression of the B gene from a larger transcript was also predicted, since no identifiable transcription terminator existed between the tail gene D and B (Section 6.1).

Northern analysis was used to identify the 290 b B gene transcript *in vivo* (Section 6.2.1(a)) and the activity of the pB promoter *in vivo* was confirmed by sequencing the 5'-end of the 290 b transcript (Section 6.2.1(b)). Northern analysis revealed that, late in infection, the B gene was transcribed onto a larger transcript sized at 2.0 kb, and a preliminary experiment indicated that this transcript initiated to the left of the HindIII site at 61.3% (Section 6.2.1(a)). It is pertinent to note that the late genes immediately to the left of gene B comprise the GFED polarity group (Hocking and Egan, 1982b), and gene B is therefore probably represented on the GFED late transcript. Since the B gene product controls late gene expression at the level of transcription (Finnegan and Egan, 1981), it is probable that this late promoter is under positive control of the B protein. Gene G was mapped to the immediate left of the BamHI site at 58.7% (Finnegan and Egan, 1979) and the minimum size of a transcript encoding genes G through B, and terminating at terminator tB (67.4%), is 2.6 kb. This is considerably larger than the 2.0 kb transcript detected by Northern analysis and further characterisation of the larger transcript from which the B gene is expressed, awaits DNA sequencing of the late region to the left of the B gene.

The P2 late control gene ogr, like the 186 B gene, is transcribed *in vitro* to yield a 290 b transcript (Pritchard, 1984; Pritchard and Egan, 1985) and the ogr gene may also be expressed from a larger transcript late in infection. This was concluded from the DNA sequence of the P2 ogr promoter-regulatory region, where no potential rho-independent transcription terminator(s) could be recognised between the reading frame of the presumptive P2 tail gene D and the late control gene ogr (Section 7.4.3).

9.3.2 Expression of the B Gene and Control of Late Gene Transcription

The 290 b B gene transcript was shown to be the strongest *in vitro* transcript (Pritchard, 1984; Pritchard and Egan, 1985). This result was unexpected since earlier studies provided evidence that transcription of the B gene was dependent upon phage protein synthesis (Finnegan and Egan, 1981). Transcription of the B gene *in vitro* by unmodified *E. coli* RNA polymerase holoenzyme led to the prediction that the B gene transcript would be present in the lysogen since there is no evidence of immunity determinants to the left of the attachment site *att* (the virulent mutant *vir2* was mapped to the right of the *att-int* region, Hocking and Egan, 1982a).

Northern and primer extension analysis was used to confirm that the 290 b B gene transcript was indeed transcribed in the lysogen (Section 7.4.1, 7.4.2). Evidence was provided that transcription of the B gene was controlled by the B protein itself. This was concluded from the marked increase in the amount of the 290 b B gene transcript when the prophage carried a defective B gene (Section 7.4.1, 7.4.2). B gene transcription was therefore predicted to be under repressive autogenous control.

Finnegan and Egan (1981) provided evidence that the B gene product acts positively to control late gene transcription and since the B gene is transcribed from the prophage, it would be expected to activate late gene transcription. Late gene expression from the prophage does not occur, as indicated by the marker rescue frequencies for 186B^{+}*vir* late *am* mutants (late genes D-W) from a 186*cItsp* prophage, which were consistent with marker rescue by recombination rather than by complementation (Hocking and Egan, 1982a). It is important to note that the failure to activate late genes from the prophage does not result from the inability of the B protein to act in *trans*, since the B protein from the prophage

can act in *trans* to turn on the late genes of a superinfecting *Bam* phage (Hocking and Egan, 1982a).

The inability to activate late gene expression from the prophage is therefore predicted to reflect the requirement of a *cis*-acting function which is necessary in addition to the *B* gene product, to activate 186 late gene expression. This *cis*-acting function may be some as yet undefined gene-product(s), presumably under cI repressor control in the prophage, or the product of gene *A* which has been shown to act in *cis* (Hocking, 1977), or a replicating template.

Hybridisation studies carried out by Finnegan and Egan (1981) indicated that late gene transcription does not occur after heat induction of a *AamcItsp* prophage and they suggested either a direct or indirect involvement of the A protein in the control of late gene transcription. Some evidence does exist that the *A* gene product itself is not required for *B*-dependent activation of late gene expression. This comes from a phasmid (a 186 phage-plasmid recombinant) which was constructed by H. Richardson in our laboratory. The 3.6 kb *XhoI-BglII*(67.6%-79.6%) fragment from a 186*AamcItsp* phage was replaced by the 4.5 kb *XhoI-BglII* fragment containing the origin of replication of the plasmid pKC7 (Nagaraja and Rogers, 1979). This phasmid forms plaques with the same efficiency on both suppressing and non-suppressing bacterial strains, suggesting that replication of the phasmid was from the plasmid origin of replication in the non-suppressing strain (H. Richardson, personal communication). The ability to express late genes from the phasmid, in the absence of functional 186 A protein, could reflect an indirect rather than a direct need for the A protein during the normal course of 186 late gene expression. Alternatively, as described previously, the possibility exists that an undefined *cis*-acting 186 gene-product

is required for B-dependent activation of late genes. This cis-acting gene cannot be located in the XhoI-BglII(67.6%-79.6%) region of phage 186 since this region was deleted in the phasmid.

Evidence has been presented above that a replicating template rather than the product of replication gene A is required for B-dependent activation of late gene transcription. The requirement for a replicating template could simply be to increase the gene dosage of an undefined cis-acting gene-product(s) to a level required for late gene expression. A more interesting possibility is that the replicating template provides a favourable DNA conformation which is necessary for the B protein to effect positive control of late gene transcription. Menzel and Gellert (1983) provided evidence that the E. coli gyrase genes are regulated in a manner dependent on DNA conformation.

To this point, evidence has been presented that a replicating template or an undefined cis-acting 186 gene-product(s) is required for B-dependent activation of 186 late gene expression. Replication seems to play an important role in the control of B gene transcription itself. Finnegan and Egan (1981) noted the markedly reduced transcriptional activity of the pEC35 region (Fig. 1.3), which was presumed to reflect B gene transcription after heat induction of a AamcItsp lysogen (See also Section 1.5.2). The conclusion drawn from these results was that B gene transcription was either directly or indirectly dependent on the activity of the A gene-product. It seems likely that a replicating template is required for B gene transcription rather than the A protein itself since the phasmid described previously does not contain functional A protein when grown on a non-suppressing bacterial strain, and yet sufficient B protein is produced to allow expression of late genes.

To summarise, a replicating template is predicted to be needed, in addition to B protein, for the activation of 186 late

gene transcription, and also for full induction of B gene transcription.

The existence of repressive autogenous control for B gene transcription, and the need of a replicating template for B action, would be properties consistent with the need of replication gene A activity for the induction of B gene transcription and for the expression of the late genes, as reported by Finnegan and Egan (1981). Replication which is controlled by the A gene product (Hocking and Egan, 1982c) provides the replicating template presumed to be essential for B-dependent activation of late gene transcription and, based on the number of polarity groups found in the late region (Hocking and Egan, 1982b), it is expected that there will be at least four sites on each replicating template where B-dependent activation can occur. Replication also leads to an increase in template numbers. These two consequences of replication would lead to the titration of a finite pool of B protein originally present in the lysogen (resulting from autogenous negative control of B gene transcription) and thereby induction of B gene transcription. Additionally, late in infection the B gene appears on a larger transcript (Section 6.2.1(a)), which may further enhance the level of the B gene product when its activity is required most.

Repressive autogenous control of B gene transcription and the need of a replicating template would explain the very high efficiency of plating when immunity-insensitive (vir) Bam mutants of phage 186 were plated on non-supressing 186 lysogens, sup^{o}(186) (Hocking and Egan, 1982a). The explanation now offered for these results is that the superinfecting virBam phage replicates, and the increase in template numbers titrates the available B protein from the prophage, which induces B gene transcription from the prophage and in turn leads to activation of late gene transcription from the replicating superinfecting phage.

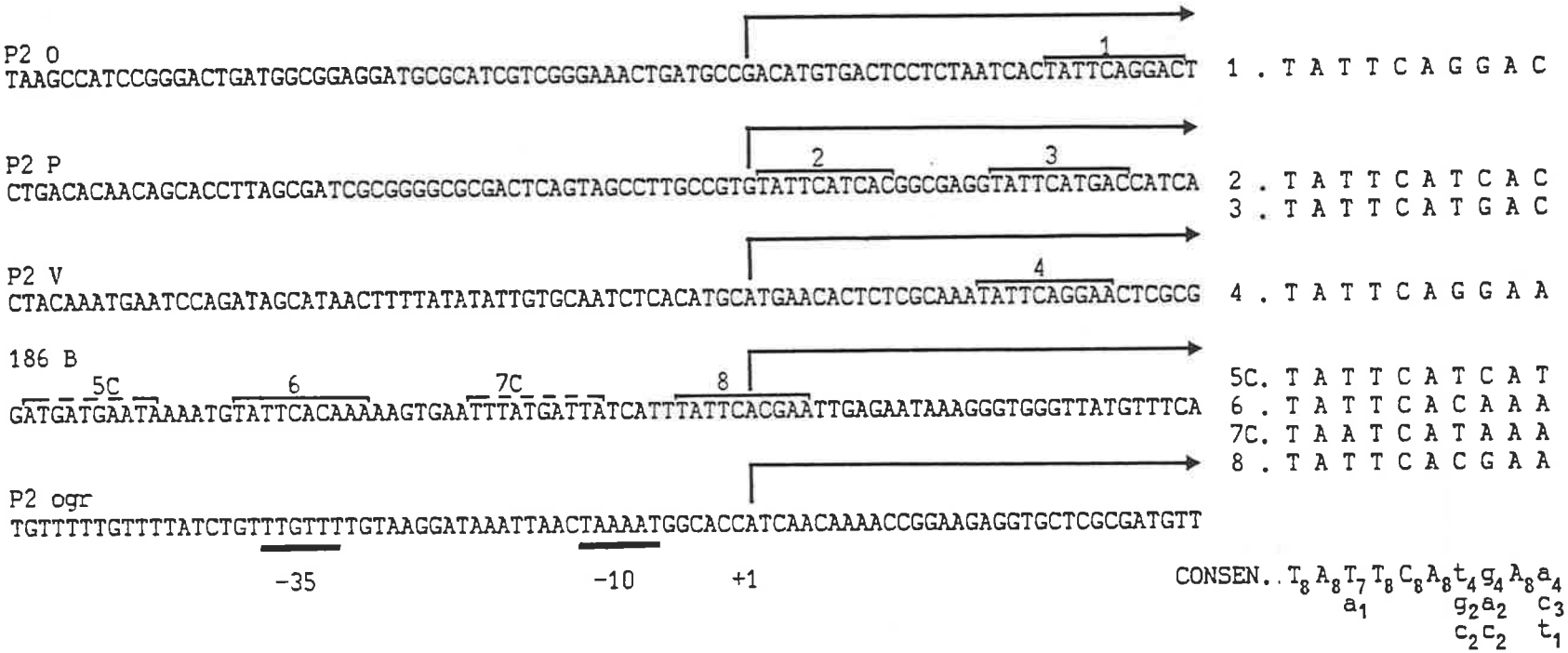
9.3.3 The B Box

The predicted properties of the B protein are that it interacts with the DNA at the B promoter to prevent transcription, and with the DNA at the late promoter(s) to activate transcription. A comparison of the pB promoter with the 186 late promoter(s) may reveal conserved DNA sequences important for B action, however the sequence(s) of the 186 late promoter(s) is not known. The 5'-ends of four P2 late mRNAs have been identified and the DNA sequence of the promoter region for each of these transcripts has been determined (Christie and Calendar, 1983, 1985). The proposed promoter regions show no significant homology to the E.coli consensus promoter, as would be expected for positively controlled promoters, but were found to have homology among themselves. A consensus sequence which may be recognised by a modified E. coli RNA polymerase was proposed for these promoters. Hybrid phages (Hy 9, 10, 11 - Hocking and Egan, 1982d) have been constructed where P2 late gene expression is under 186 B protein control and therefore it was predicted that the 186 B protein would also positively control P2 late promoters.

As described previously (Sections 4.2.6, 7.3), two 20 bp inverted repeats were detected spanning the -10 and -35 regions of the 186 pB promoter. It was provocative to find the 10 bp arm of each repeat corresponds to the highly conserved 10 bp sequence found by Christie and Calendar (1983, 1985) at around +20 in three of the four P2 late promoter regions (Fig. 9.2). This 10 bp sequence has been termed the B box. Binding of the B protein to the B box at the -10 and -35 regions of the pB promoter would block access of the RNA polymerase to the promoter and effect repression, whilst binding at the +20 region of the late promoter(s) could facilitate RNA polymerase binding by some as yet unknown mechanism.

Fig. 9.2 The B Box

Promoter regions of the P2 genes O, P and V (Christie and Calendar, 1983, 1985) and the promoters of the 186 B and P2 ogr genes are presented aligned at the proposed mRNA start point of each gene (+1). The -10 and -35 regions of the 186 B and P2 ogr promoters are underlined. Boxed regions 1, 2, 3, 4, 6 and 8 are found in the l-strand of the sequence and boxes 5C and 7C are found in the complementary strand. The consensus sequence for the proposed B box is shown on the right.


P2 O
TAAGCCATCCGGGACTGATGGCGGAGGATGCGCATCGTCGGGAAACTGATGCCGACATGTGACTCCTCTAATCACTATTCAGGACT
1
1 . T A T T C A G G A C
P2 P
CTGACACAACAGCACCTTAGCGATCGCGGGGCGCGACTCAGTAGCCTTGCCGTGTATTCATCACGGCGAGGTATTCATGACCATCA
2
3
2 . T A T T C A T C A C
3 . T A T T C A T G A C
P2 V
CTACAAATGAATCCAGATAGCATAACTTTTATATATTGTGCAATCTCACATGCATGAACACTCTCGCAAATATTCAGGAACTCGCG
4
4 . T A T T C A G G A A
186 B
GATGATGAATAAAATGTATTCACAAAAAGTGAATTTATGATTATCATTTATTCACGAATTGAGAATAAAGGGTGGGTTATGTTTCA
5C
6
7C
8
5C. T A T T C A T C A T
6 . T A T T C A C A A A
7C. T A A T C A T A A A
8 . T A T T C A C G A A
P2 ogr
TGTTTTTGTTTTATCTGTTTGTTTTGTAAGGATAAATTAACTAAAATGGCACCATCAACAAAACCGGAAGAGGTGCTCGCGATGTT
-35
-10
+1
CONSEN.. T8 A8 T7 T8 C8 A8 t4 g4 A8 a4
a1
g2 a2 c3
c2 c2 t1

The transcription of the P2 ogr gene in vitro in the absence of phage protein synthesis (Pritchard, 1984; Pritchard and Egan, 1985) suggests that in vivo the P2 ogr gene should, like the 186 B gene, be transcribed early after infection and in the lysogen. Therefore some control must be operative to prevent induction of late gene transcription by the ogr gene product. DNA sequence analysis of the P2 ogr promoter-regulatory region (Fig. 7.6) did not result in the detection of a B box in the promoter region of the P2 ogr gene; should the B box be recognised by the P2 late control protein Ogr at the P2 late promoters, then it would suggest that the late control gene of P2, in contrast to the situation found in 186, does not negatively control its own transcription.

Any mechanism which accounts for the control of ogr transcription and the activation of P2 late gene transcription by Ogr, must take into account the requirement for the P2 replication genes A and B in the activation of the P2 late genes (Lindahl, 1970; Lindqvist, 1971). In vivo studies such as those used to characterise 186 B gene transcription are required to investigate the control of P2 ogr gene transcription.

9.3.4 Comparison of Late Gene Transcription in Phages 186, P2 and λ

Late gene transcription in phages 186 and P2 (members of the P2-related family) is very different to that found in the lambdoid phages. The λ Q protein acts as a positive regulator of late gene transcription by causing antitermination of the transcript initiating at the pR' promoter, which then becomes a leader for the λ late gene transcript (Roberts, 1975).

Evidence has been presented that the 186 B protein is also a positive regulator of late gene transcription (Finnegan and Egan,

1981), but evidence has been presented that the B protein is also a negative regulator of its own transcription. Late gene transcription in P2 is positively controlled by the Ogr protein (Sunshine and Sauer, 1975) and late gene expression is predicted to require the activation of at least four positively controlled late promoters (Christie and Calendar, 1983, 1985) rather than antitermination. As the 186 B protein is capable of substituting for P2 Ogr in hybrid P2.186 phages (Hy 9, 10, 11 - Hocking and Egan, 1982d), a similar control of 186 late gene expression is expected.

The P2 Ogr protein mediates the turn-on of P2 late gene expression through an interaction with the α-subunit of E. coli RNA polymerase (Fujicki et al., 1976). Since the P2 ogr and 186 B genes are functionally interchangeable and show extensive DNA and amino acid sequence homology (Section 8.7), a similar interaction with the α-subunit of E. coli RNA polymerase might be expected for the 186 B protein. No specific role has been assigned for the α-subunit of E. coli RNA polymerase and it is rarely implicated as having a role in selective transcription. Studies intended to further the understanding of the control of late gene expression in the P2-related phages may also provide insights into the function of the α-subunit of E. coli RNA polymerase.

9.3.5 Future Work

Studies of the control of 186 B gene transcription presented in this work provided evidence for repressive autogenous control of B gene transcription. This proposal can be tested directly by addition of B protein to an in vitro transcription reaction which should interfere with the synthesis of the 290 b B gene transcript. The binding site of the B protein on 186 DNA can further be defined by DNA protection experiments similar to those described by Schmitz and Galas (1979). Both the above studies require isolation of the

B protein. The B gene can be cloned into the high expression vector of Remaut et al. (1983) and the B protein isolated by standard protein purification techniques. Site-directed mutagenesis (Zoller and Smith, 1983) could be used to determine the importance of the B boxes overlapping the pB promoter by altering the highly conserved bases in the B box and investigating the effects of such a change on B gene transcription.

Identification of the 186 late promoter(s) will require the sequence of the late region of 186 to be determined. This will allow the construction of defined DNA probes to detect the late transcript(s), and DNA primers to locate the 5'-ends of the late transcripts by primer extension. Late promoters can then be identified and examined for the existence of the B box. These studies are currently underway in this laboratory.

Finally, in order to define the minimal requirements for late gene expression (other than the B gene and late genes) in phage 186, the phasmid previously described (Section 9.3.2) can be used. Further deletions to the right of the BglII(79.6%) site will determine if any additional genes are required for late gene expression and site-specific mutagenesis can be used to identify the gene(s).

SECTION 10

PUBLICATIONS AND PRESENTATIONS AT MEETINGS

10. PUBLICATIONS AND PRESENTATIONS AT MEETINGS

Kalionis, B. and Egan, J.B. (1981). Orientation of separated DNA strands of coliphage 186 relative to its genetic map. Gene, 15, 95-98.

Kalionis, B., Pritchard, M. and Egan, J.B. (1985). Control of gene expression in the P2-related phages. IV. The late control gene B of coliphage 186. Submitted J. Mol. Biol.

Kalionis, B., Pritchard, M. and Egan, J.B. (1985). Control of gene expression in the P2-related phage. Comparison of the late control proteins of coliphages P2, P4 and 186. Submitted J. Mol. Biol.

Kalionis, B., Dodd, I.B. and Egan, J.B. (1985). Control of gene expression in the P2-related phages. III. DNA sequence of the early control region of coliphage 186. Submitted J. Mol. Biol.

Richardson, H.E., Kalionis, B. and Egan, J.B. Control of gene expression in the P2-related phages VI. Control of middle gene transcription of coliphage 186. Manuscript in preparation.

Kalionis, B. and Egan, J.B. (1980). Orienting 186 DNA strands relative to its genetic map. Proceedings of the Australian Biochemical Society, 13, 88.

Kalionis, B., Pritchard, M., Dodd, I., Sivaprasad, A.V., Puspurs, A. and Egan, J.B. (1984). Control of gene expression in temperate coliphage 186. Organisation and Expression of the Eukaryotic Genome. 6th Annual Conference, Lorne, Victoria.

Kalionis, B., Pritchard, M. and Egan, J.B. (1985). Control of late gene expression in coliphage 186. Organisation and Expression of the Genome. 7th Annual Conference, Lorne, Victoria.

Dodd, I., Kalionis, B. and Egan, J.B. (1985). Genetic switching mechanisms in the template coliphages λ and 186. Organisation and Expression of the Genome. 7th Annual Conference, Lorne, Victoria.

SECTION 11

BIBLIOGRAPHY

Adhya, S. and Gottesman, M. (1978). Annu. Rev. Biochem. 47, 967-996.

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

Akusjarvi, G. and Pettersson, U. (1979). Cell, 16, 841-850.

Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982). Cell, 29, 319-328.

Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B. (1981). Nature (London), 290, 754-758.

Appelyard, R.K. (1954). Genetics, 39, 440-452.

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

Barrell, B.G. (1971). In: Procedures in Nucleic Acid Research (Eds. Cantoni, G.L. and Davies, D.R.) Harper and Row, N.Y. pp. 751-779.

Benzer, S. (1955). Proc. Nat. Acad. Sci., U.S.A. 41, 344-354.

Berk, A.J. and Sharp, P.A. (1977). Cell, 12, 721-732.

Bertani, G. (1968). In: Molecular Genetics. (Eds. Wittman, H. and Schuster, N.) Springer-Verlag, Berlin, pp. 180-186.

Bertani, G., Bertani, L.E., Ljungquist, E. and Westöo, A. (1982). Genet. Maps, 2, 43-47.

Bertani, L.E. (1968). Virology, 36, 87-103.

Bertani, L.E. (1970). Proc. Nat. Acad. Sci., U.S.A. 65, 331-336.

Bertani, L.E. (1976). Virology, 71, 85-96.

Bertani, L.E. and Bertani, G. (1971). Adv. Genet. 16, 199-237.

Better, M., Wickner, S., Auerbach, J. and Echols, H. (1983). Cell, 32, 161-168.

Birmboin, H.C. and Doly, J. (1979). Nucl. Acids Res. 7, 1513-1523.

Bøvre, K. and Szybalski, W. (1971). Meths. in Enzymology, 21, 350-383.

Bradley, C., Ong, P.L. and Egan, J.B. (1975). Mol. gen. Genet. 140, 123-135.

Brendel, V. and Trifonov, E.N. (1984). Nucl. Acids Res. 12, 4412-4427.

Brosius, J., Dull, T.J. and Noller, H.F. (1980). Proc. Nat. Acad. Sci., U.S.A. 77, 201-204.

Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978). Proc. Nat. Acad. Sci., U.S.A. 75, 4801-4805.

Bruening, G., Gould, A.R., Murphy, P.J. and Symons, R.H. (1982). F.E.B.S. Letters, 148, 71-78.

Campbell, A. (1962). Adv. Genet. 11, 101-116.

Campbell, A. (1965). Virology, 27, 329-339.

Capaldi, R.A. and Vanderkooi, G. (1972). Proc. Nat. Acad. Sci., U.S.A. 69, 930-932.

Chattoraj, D.K. (1978). Proc. Nat. Acad. Sci., U.S.A. 75, 1685-1689.

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

Chattoraj, D.K. and Bertani, G. (1980). Mol. gen. Genet. 178, 85-90.

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

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

Choe, B.K. (1969). Mol. gen. Genet. 105, 275-284.

Christie, G.E. and Calendar, R. (1983). J. Mol. Biol. 167, 773-790.

Christie, G.E. and Calendar, R. (1985). J. Mol. Biol. 181, 373-382.

Court, D., Brady, C., Rosenberg, M., Wulff, D.L., Behr, M., Mahoney, M. and Izumi, S. (1980). J. Mol. Biol. 138, 231-254.

Court, D.H., Huang, T.F. and Oppenheim, A.B. (1983). J. Mol. Biol. 166, 233-240.

Craig, N. and Nash, H.A. (1983). Cell, 35, 795-803.

Dambly-Chaudière, C., Gottesman, M., Debouck, C. and Adhya, S. (1983). J. of Mol. and Appld. Gen. 2, 45-56.

Daniels, D.L., Sanger, F. and Coulson, A.R. (1982). Cold Spring Harbor Symp. Quant. Biol. 47, 1009-1024.

Davis, R.W. and Parkinson, J.S. (1971). J. Mol. Biol. 56, 403-423.

Dharmarajah, V.K. (1975). Honours Thesis, Adelaide University.

Diana, C., Deho, G., Geisselsoder, J., Tinelli, L. and Goldstein, R. (1978). J. Mol. Biol. 126, 433-445.

Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977). Nucl. Acids Res. 4, 2527-2538.

Echols, H. (1980). In: The Molecular Genetics of Development (Eds. Loomis, W. and Leighton, T.) Academic, N.Y., pp. 1-16.

Echols, H., Green, L., Oppenheim, A.B., Oppenheim, A. and Honigman, A. (1973). J. Mol. Biol. 80, 203-216.

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

Fickett, J.W. (1982). Nucl. Acids Res. 10, 5303-5318.

Finnegan, J. (1979). Ph. D. Thesis, University of Adelaide.

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

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

Forbes, D. and Herskowitz, I. (1982). J. Mol. Biol. 160, 549-569.

Friedman, D.I., Olson, E.R., Georgopoulos, C., Tilly, K., Herskowitz, I. and Banuett, F. (1984). Microbiol. Rev. 48, 299-325.

Friedman, D.I. and Olson, E.R. (1983). Cell, 34, 143-149.

Funnell, B.E. and Inman, R.B. (1982). J. Mol. Biol. 154, 85-101.

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

Fujicki, H., Palm, P., Zillig, W., Calendar, R. and Sunshine, M. (1976). Mol. gen. Genet, 145, 19-22.

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

Geisselsoder, J., Mandel, M., Calendar, R. and Chattoraj, D.K. (1973). J. Mol. Biol. 77, 405-415.

Geisselsoder, J., Youderian, P., Deho, G., Chidambaram, M., Goldstein, R. and Ljungquist, E. (1981). J. Mol. Biol. 148, 1-19.

Ghisotti D., Zangrossi, S. and Sironi, G. (1983). J. Virol. 48, 616-626.

Gicquel-Sanzey, B. and Cossart, P. (1982). EMBO J. 1, 591-595.

Gillespie, D. and Spiegelman, S. (1965). J. Mol. Biol. 12, 829-842.

Glisin, V., Crkvenjakov, R. and Byus, C. (1974). Biochemistry, 13, 2633-2637.

Godson, G.N. (1980). Fed. Procs. 39, 2822-2829.

Goldstein, R., Sedivy, J. and Ljungquist, E. (1982). Proc. Nat. Acad. Sci., U.S.A. 79, 515-519.

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

Goosen, N., van Heuvel, M., Moolenaar, G.F. and van de Putte, P. (1974). Gene, 32, 419-426.

Gottesman, M.E. and Yarmolinsky, M.B. (1968). J. Mol. Biol. 31, 487-505.

Grosjean, H. and Fiers, W. (1982). Gene, 18, 199-209.

Grosschedl, R. and Hobom, G. (1979). Nature (London), 277, 621-627.

Guarneros, G., Montanez, C., Hernandes, T. and Court, D. (1982). Proc. Nat. Acad. Sci., U.S.A. 79, 238-242.

Guarneros, G. and Echols, H. (1970). J. Mol. Biol. 47, 565-583.

Hall B.D., Haarr, L. and Kleppe, K. (1980). TIBS, 5, 254-256.

Hawley, D.K. and McClure, W.R. (1983). Nucl. Acids Res. 11, 2237-2255.

Hershey, A.D., Burgi, E. and Davern, C.I. (1965). Biochem. Biophys. Res. Comm. 18, 675-678.

Herskowitz, I. and Hagen, D. (1980). Annu. Rev. Genet. 14, 399-445.

Ho, Y-S., Wulff, D. and Rosenberg, M. (1983). Nature (London), 304, 703-708.

Ho, Y-S. and Rosenberg, M. (1982). Ann. Microbiol. (Paris). 133A, 215-218

Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwarz, E. and Kossel, H. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 165-178.

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

Hocking, S.M. and Egan, J.B. (1982a). Mol. gen. Genet. 187, 87-95.

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

Hocking, S.M. and Egan, J.B. (1982c). J. Virol. 44, 1068-1071.

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

Hoess, R.H., Foeller, C., Bidwell, K. and Landy. A. (1980). Proc. Nat. Acad. Sci., U.S.A. 77, 2482-2486.

Hoess, R.H. and Landy, A. (1978). Proc. Nat. Acad. Sci., U.S.A. 75, 5437-5441.

Holmes, W.M., Platt, T. and Rosenberg, M. (1983). Cell, 32, 1029-1032.

Hooper, I. (1979). PhD. Thesis, University of Adelaide.

Horii, T., Ogawa, T., Nakatani, T., Hase, T., Matsubara, H. and Ogawa, H. (1981). Cell, 27, 515-522.

Howard, B. (1967). Science, 158, 1588-1589.

Humayun, Z., Jeffrey, A. and Ptashne, M. (1977). J. Mol. Biol. 112, 265-277.

Inman, R.B., Schnos, M., Simon, L.D. Six, E.W. and Walker, D.H. (1971). Virology, 44, 67-72.

Johnson, A., Meyer, B.J. and Ptashne, M. (1979). Proc. Nat. Acad. Sci., U.S.A. 76, 5061-5065.

Johnson, A.D., Pabo, C.O. and Sauer, R.T. (1980). Meths. in Enzymology. 65, 839-856.

Johnson, A.D., Poteete, A.R., Lauer, G. Sauer, R.T., Ackers, G.K. and Ptashne, M. (1981). Nature (London), 294, 217-223.

Jones, I.M., Primrose, S.B. and Ehrlich, S.D. (1982). Mol. gen. Genet. 188, 486-489.

Kalionis, B., Pritchard, M. and Egan J.B. (1985). J. Mol. Biol. Submitted.

Kalionis, B. and Egan, J.B. (1981). Gene, 15, 95-98.

Kourilsky, P., Marcaud, L., Sheldrick, P., Luzzatti, D. and Gros, F. (1968). Biochemistry, 61, 1013-1020.

Leong, J.M., Nunes-Duby, S., Lesser, C.F., Youderian, P., Susskind, M.M. and Landy, A. (1985). J. Biol. Chem. 260, 4468-4477.

Landsmann, J., Kroger, M. and Hobom, G. (1982). Gene, 20, 11-24.

Landy, A. and Ross, W. (1977). Science, 197, 1147-1160.

Laughon, A. and Scott, M.P. (1984). Nature (London), 310, 25-31.

Lin, C.S. (1984). Nucl. Acids Res. 12, 8667-8684.

Lindahl, G. (1969). Virology, 39, 839-860.

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

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

Lindahl, G., Sironi, G., Bialy, H. and Calendar, R. (1970). Proc. Nat. Acad. Sci., U.S.A. 66, 587-594.

Lindahl, G. and Sunshine, M. (1972). Virology, 49, 180-187.

Lindqvist, B.H. (1971). Mol. gen. Genet. 110, 178-196.

Lindqvist, B.H. (1974). Proc. Nat. Acad. Sci., U.S.A. 71, 2752-2755.

Lindqvist, B.H. and Bøvre, K. (1972). Virology, 49, 690-699.

Lindqvist, B.H. and Six, E. (1971). Virology, 43, 1-7.

Little, J.W., Edmiston, S.H., Pacelli, L.Z. and Mount, D.W. (1980). Proc. Nat. Acad. Sci., U.S.A. 77, 3225-3229.

Little, J.W. and Mount, D.W. (1982). Cell, 29, 11-22.

Ljungquist, E., Kockum, K. and Bertani, L.E. (1984). Proc. Nat. Acad. Sci., U.S.A. 81, 3988-3992.

Ljungquist, E. and Bertani, L.E. (1983). Mol. gen. Genet. 192, 87-94.

Lundqvist, B. and Bertani, G. (1984). J. Mol. Biol. 178, 629-651.

Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A. and Maurer, R. (1975) Cell, 5, 109-113.

Mascarenhas, D., Trueheart, J., Benedik, M. and Campbell, A. (1983). Virology, 124, 100-108.

Matthews, B.W., Grutter, M.G., Anderson, W.F. and Remington, S.J. (1981). Nature (London), 290, 334-335.

Matthews, B.W., Ohlendorf, D.H., Anderson, W.F. and Takeda, Y. (1982). Proc. Nat. Acad. Sci., U.S.A. 79, 1428-1432.

Maxam, A.M. and Gilbert,W. (1977). Proc. Nat. Acad. Sci., U.S.A. 74, 560-564.

Maxam, A.M. and Gilbert, W. (1980). Meths. in Enzymology, 65, 499-560.

McKay, D.B. and Steitz, T.A. (1981). Nature (London), 290, 744-749.

McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. (1981). In: Gene Amplification and Analysis, Vol. 2, Structural Analysis of Nucleic Acids (Eds. Chrikjian, J.G. and Papas, T.S.) Elsevier, N.Y. pp. 383-415.

McKnight, S.L., Gavis, E.R. and Kingsbury, R. (1981). Cell, 25, 385-398.

McMacken, R., Mantei, N., Butler, B., Joyner, A. and Echols, H. (1970). J. Mol. Biol. 49, 639-655.

McMaster, G.K. and Carmichael, G.G. (1977). Proc. Nat. Acad. Sci., U.S.A. 74, 4835-4838.

Menzel, R. and Gellert, M. (1983). Cell, 34, 105-113.

Messing, J. (1979). Recombinant DNA Technical Bulletin, 2, 43-48.

Messing, J. (1983). Meths. in Enzymology. 101, 20-78.

Messing, J., Crea, R. and Seeburg, P.H. (1981). Nucl. Acids Res. 9, 309-323.

Messing, J., Gronenborn, B., Müller-Hill, B. and Hofschneider P.H. (1977). Proc. Nat. Acad. Sci., U.S.A. 74, 3642-3646.

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

Meyer, B.J. and Ptashne, M. (1980). J. Mol. Biol. 139, 195-205.

Miller, H.I., Abraham, J., Benedik, M., Campbell, A., Court, D., Echols, H., Fischer, R., Galindo, J.M., Guarneros, G., Hernandez, T., Mascarenhas, D., Montanez, C., Schindler, D., Schmeissner, U. and Sosa, L. (1981). Cold Spring Harbor Symp. Quant. Biol. 45, 439-445.

Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., Hsu, P.-L., Ross, W. and Landy, A. (1981). Cold Spring Harbor Symp. Quant. Biol. 45, 429-437.

Moore, D.D., Denniston-Thompson, K., Kruger, K.E., Furth, M.E., Williams, B.G., Daniels, D.L. and Blattner, F.R. (1979). Cold Spring Harbor Symp. Quant. Biol. 43, 155-163.

Mulligan, M.E., Hawley, D.K., Entriken, R. and McClure, W.R. (1984). Nucl. Acids Res. 12, 789-800.

Murray, K., Isaksson-Forsen, G., Challberg, M. and Englund, P.T. (1977). J. Mol. Biol. 112, 471-489.

Murray, K. and Murray, N.E. (1973). Nature (New Biology), 243, 134-139.

Nagaraja, R. and Rogers, S.G. (1979). Gene, 7, 79-82.

Nilsson, E. and Bertani, L.E. (1977). Mol. gen. Genet. 156, 297-302.

Nobrega, F.G., Dieckmann, C.L. and Tzagoloff, A. (1983). Anal. Biochem. 131, 141-145.

Nygaard, A.P. and Hall, B.D. (1963). Biochem and Biophys. Res. Comm. 12, 98-104.

Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y. and Matthews, B.W. (1982). Nature (London), 298, 718-723.

Oppenheim, D.S. and Yanofsky, C. (1980). Genetics, 95, 785-795.

Pabo, C.O., Sauer, R.T., Sturtevant, J.M. and Ptashne, M. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 1608-1612.

Pabo, C.O. and Lewis, M. (1982). Nature (London), 298, 443-447.

Pabo, C.O. and Sauer, R.T. (1984). Annu. Rev. Biochem. 53, 293-321.

Padmanabhan, R. and Wu, R. (1972). J. Mol. Biol. 65, 447-467.

Parkinson, J.S. and Huskey, R.J. (1971). J. Mol. Biol. 56, 369-384.

Phizicky, E. and Roberts, J. (1980). J. Mol. Biol. 139, 319-328.

Pirotta, V. (1973). Nature (New Biology), 244, 13-14.

Pirrotta, V. (1975). Nature (London), 254, 114-117.

Pritchard, M. (1984). PhD. Thesis, Adelaide University.

Pritchard, M. and Egan, J.B. (1985). EMBO J. Submitted.

Ptashne, M. (1971). In: The Bacteriophage Lambda (ed. Hershey, A.D.) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 221-237.

Ptashne, M., Backman, K., Humayan, M.Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R.T. (1976). Science, 194, 156-161

Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R. Meyer, B.J., Pabo, C.O., Roberts, T.M. and Sauer, R.T. (1980). Cell, 19, 1-11.

Remaut, E., Stanssens, P. and Fiers, W. (1981). Gene, 15, 81-93.

Remaut, E., Tsao, H. and Fiers, W. (1983). Gene, 22, 103-113.

Rezain, A. and Williams, R.H. (1984). EMBO J. Submitted.

Richardson, H.R. and Egan, J.B. (1985). Manuscript in preparation.

Roberts, J.W. (1969). Nature (London), 224, 1168-1174.

Roberts, J.W. (1975). Proc. Nat. Acad. Sci., U.S.A. 72, 3300-3304.

Roberts, J.W., Roberts, C.W. and Craig, N.L. (1978). Proc. Nat. Acad. Sci., U.S.A. 75, 4714-4718.

Rosenberg, M. Court, D., Shimatake, H., Brady, C. and Wulff, D. (1978). Nature (London) 272, 414-423.

Rosenberg, M. and Court, D. (1979). Annu. Rev. Genet. 13, 319-353.

Rupp, W.D., Wilde, C.E., Reno, D.L. and Howard-Flanders, P. (1971). J. Mol. Biol. 61, 25-44.

Saint, R.B. (1979). PhD Thesis, Adelaide University.

Saint, R.B. and Egan, J.B. (1979). Molec. gen. Genet. 171, 79-89.

Sancar, A., Hack, A.M. and Rupp, W.D. (1979). J. Bact. 137, 692-693.

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

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

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Nat. Acad. Sci., U.S.A. 74, 5463-5467.

Sauer, B., Calendar, R., Ljungquist, E., Six, E. and Sunshine, M.G. (1982). Virology, 116, 523-534.

Sauer, B., Ow, D., Ling, L. and Calendar, R. (1981). J. Mol. Biol. 145, 29-46.

Sauer, R.T., Krovatin, W., DeAnda, J., Youderian, P. and Susskind, M.M. (1983). J. Mol. Biol. 168, 699-713.

Sauer, R.T., Pabo, C.O., Meyer, B.J., Ptashne, M. and Backman, K.C. (1979). Nature (London), 279, 396-400.

Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. (1982). Nature (London), 298, 447-451.

Sauer, R.T. and Anderegg, R. (1978). Biochemistry, 17, 1092-1099.

Schindler, D. and Echols, H.E. (1981). Proc. Nat. Acad. Sci., U.S.A. 78, 4475-4479.

Schmeissner, U., Court, D., McKenney, K. and Rosenberg, M. (1981). Nature (London), 292, 173-175.

Schmeissner, U., McKenney, K., Rosenberg, M. and Court, D. (1984a). Gene, 28, 343-350.

Schmeissner, U., McKenney, K., Rosenberg, M. and Court, D. (1984b). J. Mol. Biol. 176, 39-53.

Schmeissner, U., Court, D., Shimatake, H. and Rosenberg, M. (1980). Proc. Nat. Acad. Sci., U.S.A. 77, 3191-3195.

Schmitz, A. and Galas, D. (1979). Nucl. Acids Res. 6, 111-137.

Schreier, P.H. and Cortese, R. (1979). J. Mol. Biol. 129, 169-172.

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

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

Shea, M.A. and Ackers, G.K. (1985). J. Mol. Biol. 181, 211-230.

Shepherd, J.C.W. (1981). Proc. Nat. Acad. Sci., U.S.A. 78, 1596-1600.

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

Shore, D., Deho, G., Tsipis, J. and Goldstein, R. (1978). Proc. Nat. Acad. Sci., U.S.A. 75, 400-404.

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

Sivaprasad, A.V. (1984). PhD Thesis, Adelaide University.

Six, E.W. (1975). Virology, 67, 249-263.

Six, E.W. and Klug, C. (1973). Virology, 51, 327-344.

Six, E.W. and Lindqvist, B.H. (1971). Virology, 43, 8-15.

Six, E.W. and Lindqvist, B.H. (1978). Virology, 87, 217-230.

Smith, G.E., and Summers, M.D. (1980). Anal. Biochem. 109, 123-129.

Somasekhar, G. and Szybalski, W. (1983). Gene, 26, 291-294.

Souza, L., Calendar, R., Six, E.W. and Lindqvist, B.H. (1977). Virology, 81, 81-90.

Staden, R. (1978). Nucl. Acids Res. 5, 1013-1015.

Staden, R. (1980). Nucl. Acids Res. 8, 3673-3694.

Staden, R. (1984a). Nucl. Acids Res. 12, 551-567.

Staden, R. (1984b). Nucl. Acids Res. 12, 521-539.

Staden, R. (1984c). Nucl. Acids Res. 12, 505-519.

Staden, R., and McLachlan, A.D. (1982). Nucl. Acids. Res. 10, 141-156.

Steitz, T.A., Ohlendorf, D.H., McKay, D.B., Anderson, W.F. and Matthews, B.W. (1982). Proc. Nat. Acad. Sci., U.S.A. 79, 3097-3100.

Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982). Nucl. Acids, Res. 10, 2971-2995.

Studier, F.W. (1973). J. Mol. Biol. 79, 237-248.

Studier, F.W. and Maizel, J.V. (1969). Virology, 39, 575-586.

Summers, W.C. (1969). Virology, 39, 175-182.

Sunshine, M.G. and Sauer, B. (1975). Proc. Nat. Acad. Sci., U.S.A. 72, 2770-2774.

Susskind, M.M. and Youderian, P. (1983). In: Lambda II (Eds. Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A.) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 347-363.

Thomas, P.S. (1980). Proc. Nat. Acad. Sci., U.S.A. 77, 5201-5205.

Thomas, P.S. (1983). Meths. in Enzymology, 100, 255-266.

Tomizawa, J.I. and Ogawa, T. (1967). J. Mol. Biol. 23, 247-263.

Tsurimoto, T. and Matsubara, K. (1981). Nucl. Acids Res. 9, 1789-1799.

Tsutsui, H., Fujiyama, A., Murotsu, T., and Matsubara, K. (1983). J. Bact. 155, 337-344.

Vogelstein, B. and Gillespie, D. (1977). Bioch. and Biophys. Res. Comm. 75, 1127-1132

Wang, J.C., Martin, K.V. and Calendar, R. (1973). Biochemistry, 12, 2119-2123

Wang, J.C. and Schwartz, H. (1967). Biopolymers, 5, 953-966.

Ward, D.F. and Gottesman, M.E. (1982). Science, 216, 946-951.

Ward, D.F. and Murray, N.E. (1979). J. Mol. Biol. 133, 249-266.

Weisberg, R.A. and Gottesman, M.E. (1971). In: The Bacteriophage Lambda. (Ed. Hershey, A.D.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 489-500.

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

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

Wulff, D. and Rosenberg, M. (1983). In: Lambda II. (Eds. Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A.) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., pp. 53-73.

Younghusband, H.B., Egan, J.B. and Inman, R.B. (1975). Molec. gen. Genet. 140, 101-110.

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

Zoller, M.J. and Smith, M. (1983). Meths. in Enzymology, 100, 468.