TRANSCRIPTION TERMINATORS IN COLIPHAGE T7

by Ian Garner

A Thesis presented for the degree of

Doctor of Philosophy

Department of Molecular Biology

University of Edinburgh

Scotland

October 1982

To My Parents

ABSTRACT

Precise *in vitro* fusion of DNA fragments, in the vector system devised by McKenney and colleagues, has been used to investigate transcriptional terminators (present in the T7 genome) for the *E.coli* RNA polymerase. Three terminator-containing fragments have been studied, allowing estimation of the efficiency of each signal *in vivo* in isolation from the T7 genome. The best characterised of these is the tec1 terminator at the end of the "early" region. My estimates of a 76-78% termination of transcription at tec1 are in good agreement with the previous estimate, of Millette and colleagues, for this terminator in T7 itself. The second terminator, designated tec2 and previously studied *in vitro*, appears to be inactive *in vivo* in the tec2-fusion I have constructed. I have also identified a new transcriptional terminator of unknown physiological importance, designated tec3, from the late region of the T7 genome. I estimate a 95% efficiency of termination for this signal and have determined the DNA sequence (179 bp) of the tec3-containing T7 *AluI* fragment.

Several of these fragments stimulate termination when transcribed in the opposite direction to natural transcription of the phage genome during infection. The significance of this, and the possible involvement of a cofactor(s) interaction with *E.coli* RNA polymerase during termination at tec1 are discussed.

Acknowledgements

I would like to take this opportunity to thank all those who have made the production of this thesis possible:

Richard Hayward who has been so much more than an outstanding supervisor; Andy Newman, Chuan Ma and Kathy Howe for their stimulating companionship; all other colleagues, past and present, for their friendship, gifts and helpful discussions; Anna and Graham for their typing and photography; Alison for "brightening up" the library; and last, but not least, Lorna, for whom words fail me ....

ABBREVIATIONS

bp	base pair
Kbp	Kilobase pair
moi	multiplicity of infection
Rpol	RNA polymerase
nM	nanometres
p*X*	protein product of gene *X*
$Tet^{R/S}$	Tetracycline resistant or sensitive
$Chl^{R/S}$	Chloramphenicol resistant or sensitive
$Amp^{R/S}$	Ampicillin resistant or sensitive
att	attenuator
P*gal*	promoter for e.g. the *gal* operon

CONTENTS

CHAPTER 1 - INTRODUCTION

It is an essential requirement of all living organisms that they should be able to respond, both in the short and long term, to environmental change. The bacterium Escherichia coli can achieve this both at the primary level of gene expression, namely transcription, and at stages beyond. Here, structural genes are delineated by regulatory sequences (promoters, attenuators and terminators), and transcription is the responsibility of a single DNA-dependent RNA polymerase (Rpol). Regulation is achieved by modulating the efficiency with which E.coli Rpol can recognise and respond to these regulatory sequences.

The studies described in this thesis are an attempt to increase our knowledge of the interactions between E.coli Rpol and one type of transcriptional signal, the terminator.

COMPONENTS OF THE TRANSCRIPTION MACHINERY

Transcriptional activity in E.coli can vary drastically with physiological conditions. In accordance with this, we find that components of the system exhibit a degree of structural complexity and variability.

Rpol

A complex oligomeric protein whose properties have been extensively reviewed, for example by Losick and Chamberlin (1976), and by Lathe (1978). It consists of a catalytically competent core which interacts with a number of cofactors involved in a specific initiation or termination activities. The minimal enzyme, known as "CORE ENZYME" is composed of three distinct polypeptide subunits alpha, beta and beta-prime. These exist in the molar ratio

$\alpha_2\beta\beta'$ and the active form *in vivo* would appear to be monomeric (Zarudnaya *et al*, 1976). This pattern of $\alpha_2\beta\beta'$ in core enzyme is repeated in other prokaryotes studied (Fukuda *et al*, 1977; Burgess, 1976).

The α-subunit: This has a molecular weight of 37,000 daltons (Ovchinnikov *et al*, 1977) and is encoded by the *rpoA* gene. No clear function has yet been ascribed to α, but reconstitution of the active enzyme from purified subunits absolutely requires it (Heil and Zillig, 1970) and mutations in *rpoA* can alter transcriptional specificity (Fujiki *et al*, 1976).

The β-subunit: This is encoded by the *rpoB* gene and is a protein of 1,342 amino acid residues (Ovchinnikov *et al*, 1981). Beta has been implicated in substrate (NTP) binding (Armstrong *et al*, 1976) and genetic studies have suggested a role for it in the specificity of transcriptional initiation and termination (Wozny *et al*, 1975; Tessman and Peterson, 1976; Yamamori *et al*, 1977; Lecocq and Dambly, 1976; Guarente and Beckwith, 1977)

The β'-subunit: This is encoded by the *rpoC* gene and is a protein of 1,407 amino acid residues (Ovchinnikov *et al*, 1982). Beta-prime is a particularly basic protein (Zillig *et al*, 1970) and has been implicated in the primary binding of Rpol to DNA (Fukada and Ishihama, 1974). Other authors have suggested a role for β' in promoter selection (Gros *et al*, 1976; Coppo *et al*, 1975) although it is not predominantly determinant in this as will be seen later.

Core enzyme binds to the DNA template *in vitro* (Hinkle and Chamberlin, 1972), and catalyses the 5'→3' polymerisation of ribonucleoside 5'-triphosphates complementary to the template. $\alpha_2\beta$ precursor complex can achieve the former but not the latter step.

The reaction is effectively irreversible due to the hydrolysis and release of pyrophosphate. Efficient polymerisation by core enzyme requires the presence of a single-stranded DNA-template, a double-stranded nicked DNA-template, or a double-stranded linear DNA-template (Vogt, 1969; Burgess et al, 1969; Berg et al, 1971; Ishihama et al, 1971). Any nick can serve as a transcriptional initiation site for core enzyme (Hinkle et al, 1972). Presumably, this explains the observation that such transcription shows little or no strand specificity and initiates at random sites on the nicked templates (Hinkle and Chamberlin, 1972).

More specific Rpol activity is observed when core enzyme is complexed with another Rpol subunit, sigma (σ).

The sigma-subunit: This is an acidic protein, has a molecular weight of 82,000 daltons (Lowe et al, 1979) and is encoded by the rpoD gene. The core enzyme-sigma complex has the structure $\alpha_2\beta\beta'$-σ and is known as "HOLOENZYME". It is this form of the enzyme that will faithfully transcribe templates using true promoter sequences as transcription initiation sites. Sigma is present in the cell in less than stoichiometric amounts compared with other Rpol subunits (Ishihama et al, 1976), has a rapid turnover during transcription, dissociating from Rpol after initiation (Travers and Burgess, 1969), and appears to bind to DNA only in the $\alpha_2\beta\beta'$-σ complex (Hillel and Wu, 1978). Sigma binding to core-enzyme produces a conformational change in the enzyme structure (Wu et al, 1976) that is presumably related to the observed differences between core and holoenzyme, i.e.:

1) An altered DNA binding affinity. Core enzyme binds to many sites on DNA with an intermediate affinity, $K_{assoc} = 2\times10^{11} M^{-1}$ on the T7 genome. Holoenzyme binds to the same template resulting

in two types of complex (i) $K_{assoc} = 10^{12}$ to $10^{14} M^{-1}$, thought to be true E.coli Rpol promoters, (ii) $K_{assoc} = 10^{8}$ to $10^{9} M^{-1}$, thought to reflect a reduced general affinity of holoenzyme for DNA (Hinkle and Chamberlin, 1972).

2) An ability to melt regions of the helix in promoter:holoenzyme complexes (Chamberlin, 1976).

3) An increase in the rate of transcriptional initiation (Chamberlin, 1976).

4) A decrease in non-specific initiation at nicks. This is accompanied by an inhibition of transcription by holoenzyme due to nicks serving as tight binding sites for holoenzyme however few allow transcriptional initiation (Hinkle et al, 1972).

Sigma factor, therefore, confers transcriptional initiation specificity on the non-specific core-enzyme complex and is an essential requirement for the correct initiation of RNA synthesis.

The omega-subunit: Highly purified preparations of Rpol sometimes contain another weakly associated subunit, ω. This has a molecular weight of 8000-12000 daltons, is not essential for full activity of reconstituted Rpol (Burgess, 1976) and may, in fact, be three distinct polypeptides (Lathe, 1978).

Rpol ASSOCIATED POLYPEPTIDES OF E.coli.

Rpol may associate with a number of other cofactors during its transcriptional activities, with effects on initiation and/or termination. Although some cannot be considered true Rpol subunits, others deserve such consideration in the light of current thoughts on their activities. This discussion primarily concerns termination, but of some relevance are such associations at initiation sites. These will be mentioned briefly in the context of Rpol/cofactor interactions:

Catabolite gene activator protein (CAP)

This positive effector binds to cyclic AMP (cAMP) when the intracellular levels of the latter are high, i.e. in the absence of glucose. The resulting complex exhibits pleiotropic effects on initiation (not all positive) by binding to sequences in the promoter region of a number of operons: In the lac and ara operons, CAP-cAMP complex potentiates transcription (de Crombrugghe et al, 1971; Ogden et al, 1980). In the gal operon, however, CAP-cAMP complex stimulates initiation from one and inhibits initiation from the other of two overlapping promoters (Musso et al, 1977). These dual promoters allow a form of cellular regulation reacting to CAP-cAMP levels based on differential translation efficiency (Queen and Rosenberg, 1981). The relative efficiency of translation of the 5'-proximal and distal genes of the operon varies with the promoter used, and is therefore regulated by the CAP-cAMP complex. Sequences important for CAP-cAMP function will be discussed later, as will those relevant to other cofactors mentioned.

Repressor proteins

A number of transcription units are under negative control, two of the best understood of these are the lac and trp operons. In the former, when intracellular concentration of lactose is low, initiation is inhibited by the product of the lacI gene, the lac repressor. In this situation, repressor binds to a region overlapping the lac promoter, sterically preventing Rpol from initiating transcription (Majors, 1975a). When intracellular concentrations of lactose are high, repressor binds instead to the sugar molecule and this complex no longer interferes with initiation at the lac promoter. In the trp control region, a similar but distinct situation prevails. When intracellular concentrations of

L-tryptophan are high, the product of the trpR gene, the trp aporepressor (a 47 kdal homodimer), binds the amino acid to form an active repressor complex. This sterically blocks Rpol initiation by binding to a region overlapping the trp promoter. Aporepressor alone has a low affinity for this region and does not prevent initiation.

Repressor control of transcription also occurs in λ gene expression and the reader is referred to the review by Ptashne et al, (1980). Bacteriophage λ encodes two proteins that play contrasting roles in determining its mode of growth (see Fig. 1.1). One, the λ repressor, is the product of the cI gene and is required for lysogeny. The other, Cro, is the product of the cro gene and is required for lytic phage growth. To maintain lysogeny, repressor turns off early gene expression, including that of cro, and turns on transcription of cI. It therefore functions as both a positive and negative regulator of gene transcription. To encourage lytic growth, Cro performs two negative functions: it prevents repressor-stimulated transcription of cI; and it reduces expression (following and early burst) of early genes including cro. The effects of repressor and Cro result from their binding to two operators on the phage chromosome. The situation at the right operator is most relevant and will be described briefly. Two promoters overlap the 82 bp region defined as O_R. One of these, P_R, directs rightward transcription of a set of early genes including cro, the other, P_{RM}, directs leftward transcription of cI in a lysogen. Repressor and Cro bind to three similar 17 bp sequences in O_R (O_{R1}, O_{R2}, O_{R3}) and sterically block transcriptional initiation at P_R or P_{RM}. Binding of repressor occurs from right to left ($O_{R1} \rightarrow O_{R3}$) whereas binding of Cro occurs from left to right ($O_{R3} \rightarrow O_{R1}$). It is the

Fig. 1.1

A Partial genetic map of λ.

Promoters and terminators are boxed. The dashed arrows represent the termini and orientations of various transcripts discussed in the text. Transcripts terminate at t_{L1} and t_{R1} (and as a result of readthrough, t_{L2} and t_{R2}) unless pN antagonises termination at the four terminators shown. In the case of t_{L2}, the termination is rho-independent. The sequences of the nutL and nutR regions are shown (for the DNA strands equivalent to the mRNA).

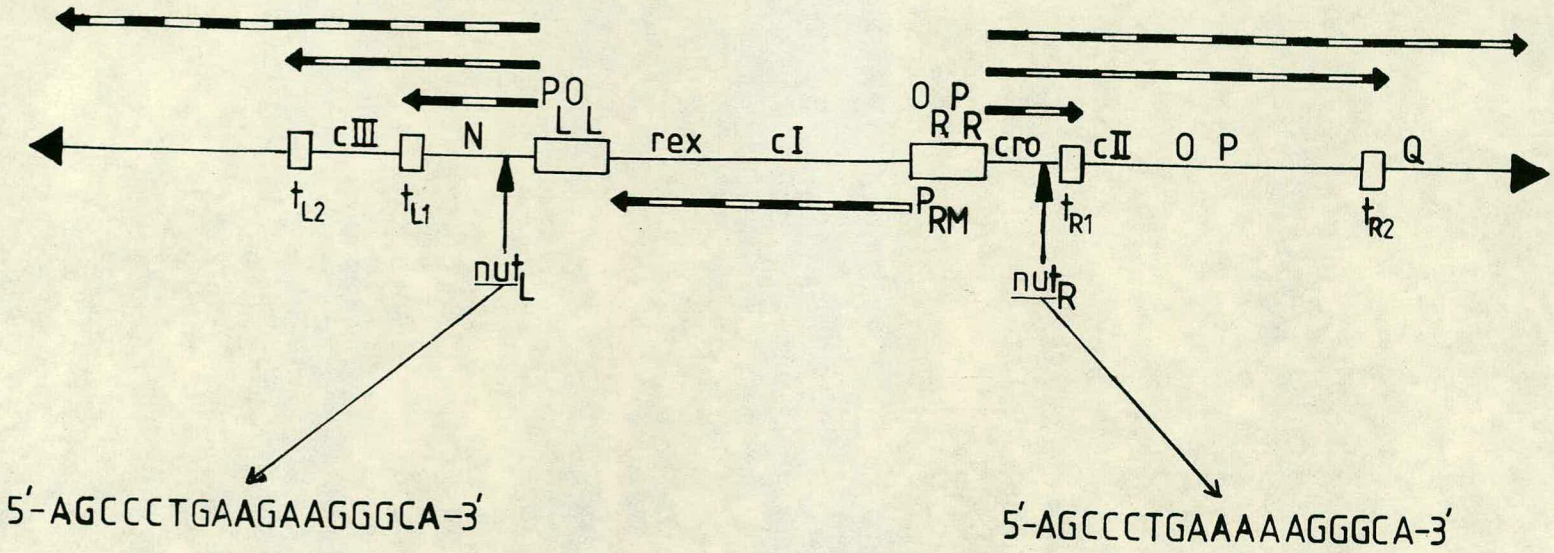

P
O
L L
cIII
N
rex
cI
O P
R R
cro
cII
O P
Q
t_{L2}
t_{L1}
P_{RM}
t_{R1}
t_{R2}
nut_L
nut_R
5'-AGCCCTGAAGAAGGGCA-3'
5'-AGCCCTGAAAAAGGGCA-3'

Fig 1·1

occupancy of these binding sites by repressor or Cro that determines which promoter is active and hence, which genes are transcribed.

Rho Protein

Rho (ρ) is a hexameric protein (Oda and Takanami, 1972; Fingel and Richardson, 1979) with a monomer weight of 50,000 daltons (Roberts, 1969; Oda and Takanami, 1972) and is the product of the rho gene (Ratner, 1976a). It exhibits two activities: (1) the termination of transcription and release of RNA at specific sites (Roberts, 1969), (2) an RNA-dependent hydrolysis of ATP and other ribonucleoside triphosphates to nucleoside diphosphates, liberating inorganic phosphate. The former activity is dependent upon the latter (Howard and de Crombrugghe, 1976). It is present in E.coli in approximately a two-fold molar excess over Rpol core complex (Blumental et al, 1976) and appears to act catalytically (Richardson, 1970) although it has been suggested that it may act stoichiometrically (Richardson and Conaway, 1980). The mechanism by which rho promotes termination remains obscure, but two lines of evidence suggest that rho and Rpol may interact. Das et al (1978) have observed that mutations in the rpoB gene restore polarity which had been relieved by a rho mutation. Moreover, the mutant Rpol terminates at the rho-dependent sites in λ only in the presence of the defective rho. This suggests a functional relationship between rho and the β-subunit of Rpol during termination. Similarly, a completed, accurate termination event at λt_{r2} and trpt' requires the presence of both pnusA and rho (Platt, 1981; Greenblatt, 1981). Greenblatt and Li (1981) have demonstrated the binding of the acidic pnusA to Rpol in vitro and proposed its association with Rpol during elongation. Rho is a basic protein and, hence, pnusA may mediate rho binding to

a paused Rpol molecule, whereupon rho would stimulate termination, possibly by causing a conformational change in the paused complex.

Other observations suggest how rho may achieve its contact with Rpol. Darlix (1973) observed that rho requires nascent RNA in order to act, termination being abolished if the RNA chain is destroyed by ribonuclease action. There is also evidence that rho binds to a specific region on the λcro mRNA *in vitro* (Bektesh and Richardson, 1980). Once attached, rho appears to entwine itself in RNA in an ATP-dependent style and can, therefore, "move" along RNA in either direction (Gallupi and Richardson, 1980). These observations led Gallupi and Richardson (op.cit.) to propose a model by which rho may achieve contact with transcribing Rpol molecules: each subunit of the rho hexamer has two binding sites for RNA and one for a nucleotide. One RNA binding site on each subunit is aligned to form a major RNA binding sites on the hexamer. The second RNA binding site on each subunit remains independent but has its affinity for RNA modulated by the state of occupancy of the nucleotide binding site in the same subunit. When ATP is bound at the nucleotide site, RNA is retained strongly by the secondary site. ATP hydrolysis causes the release of the RNA chain from this site but does not affect binding at the major site. Hence, they propose a primary binding of rho to RNA via the major site, possibly at a preferred sequence, and subsequent attachments to more distal regions via the six secondary sites. ATP hydrolysis at the nucleotide sites results in the release of RNA from secondary attachments. Subsequent ATP binding and hydrolysis allows the attachment of more distal segments of the RNA molecule to rho in either a 5' or 3' direction. In this way, rho can "wrap" RNA around itself and "migrate" along the chain in either direction

until it induces termination upon contact with an Rpol molecule. This event probably occurs at paused Rpol molecules in a site specific fashion (Shigesada and Wu, 1980) although there is evidence that rho, having contacted Rpol, may cause termination anywhere (Richardson and Conaway, 1980).

Finally, rho may, in general, act alone to mediate termination but it has recently become evident that in some cases its activities are manifested in conjunction with other cofactors (Greenblatt and Li, 1981; Ward and Gottesman, 1981). These interactions will be discussed in further detail later. However, it is presumed that some rho activity is required for cell viability as only partial rho mutants have been isolated to date.

nusA protein

This is a 69,000 dalton, acidic protein and is the product of the nusA gene located at 68' on the E.coli chromosome (Greenblatt et al, 1981). It was initially defined genetically by a mutation affecting transcription termination in bacteriophage λ (Friedman, 1971) and is identical to the 'L-factor' require for β-galactosidase synthesis in vitro (Kung et al, 1975; Greenblatt et al, 1980). Greenblatt has demonstrated the equimolar stoichiometric binding of nusA protein to Rpol core enzyme in vitro. pnusA does not bind to Rpol holoenzyme, and σ is able to displace it from the $\beta\beta'-\alpha_2$-nusA complex (Greenblatt and Li, 1981). Greenblatt proposes that pnusA associates with Rpol during elongation and that there are interchangeable, complementary forms of the enzyme as follows:

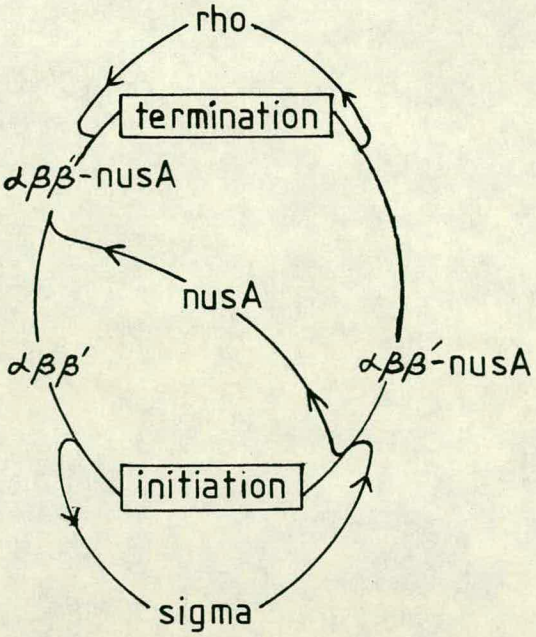


Fig1·2

Greenblatt has also demonstrated the binding of pnusA to the anti-termination factor pN of λ in vitro, but has been unable to show binding of N protein to Rpol in vitro (Greenblatt et al, 1980).

Observations in vivo and in vitro suggest that pnusA serves to mediate the interaction of Rpol with termination and antitermination factors: (i) nusA mutations relieve polarity to some degree in vivo; (ii) they prevent the growth of bacteriophage λ, as the products of both nusA and a further gene, nusB, are required for the phage N function to act (Ward and Gottesman, 1981). The requirement for nusA protein to permit N protein activity has also been demonstrated in vitro (Greenblatt et al, 1981), assayed as readthrough at the λt_{R1} and λt_{L1} terminators; (iii) nusA protein antitermination activity has been shown to be necessary for readthrough at the attenuator of the rpoBC operon in vitro (Barry et al, 1979; Greenblatt et al, 1980); (iv) termination activity of nusA protein has been demonstrated at the bacteriophage λt_{R2} terminator in vitro, where it causes Rpol to pause for 10-15 minutes, and at the trpt' terminator in vitro, where it confers specificity on the termination event. A full termination event requires the presence of rho protein (Platt, 1981).

It can clearly be seen that Rpol is influenced by several accessory proteins, probably many more than have come to light so far. Indeed, as mentioned earlier, one of those described might justifiably be promoted to the ranks of Rpol subunits, namely pnusA.

NUCLEOTIDE SEQUENCES INVOLVED IN TRANSCRIPTIONAL REGULATION

An obvious place to regulate gene expression is at the primary level of interaction between Rpol and the DNA template. Two basic mechanisms exist to achieve this: (i) variation of the nucleotide sequences involved, (ii) the use of other proteins to influence these interactions. In this section, sequences important in initiation and termination will be discussed with a bias towards the latter.

Transcriptional units in E.coli are defined by specific nucleotide sequences involved in the initiation and termination of transcription. These have been the subject of major reviews by Gilbert (1976) and Rosenberg and Court (1979). Transcription termination has been discussed more specifically by other authors (Roberts, 1976; Adhya and Gottesman, 1978a; Platt, 1981; Yanofsky, 1981).

A wide range of promoter sequences of varying efficiencies and cofactor dependencies have now been determined. Despite their functional differences, a degree of structural uniformity has emerged. The first of these is a relatively conserved 7 bp sequence centred about 10 bp upstream of the mRNA start-point (Pribnow, 1975; Schaller et al, 1975). The consensus sequence is TATPuATPu and has become known as the 'Pribnow Box' or -10 region. Promoters studied so far have at least 5 positions in agreement with this consensus sequence. Although T in the sixth position of the

heptamer is present in every case, other positions vary. Positions 1, 2 and 6 are strongly conserved, the remaining four positions being only weakly so, with A and G favoured at positions 4 and 7 respectively. The 6 bp immediately upstream of the heptamer are also conserved, but less stringently so. Here,the consensus suggested by sequence comparisons is ATTTGT, and its involvement in promoter function is supported by chemical modification contact studies (Rosenberg and Court, 1979). Hence, the conserved -10 region clearly extends for some 13 bp and the extent to which particularly positions are conserved appears to vary widely.

Sequence similarity is also notable in a region about 35 bp upstream of the mRNA start-point, known as the -35 or recognition region (Takanami et al, 1976). A highly conserved trinucleotide, TTG, occurs adjacent to and upstream of a less stringently conserved ACA sequence. This hexanucleotide occurs within a region of weaker homology resulting in a consensus sequence of TGTTGACA-TTT where positions 1, 2, 10, 11 and 12 are only weakly conserved. Position 9 has no tendency to be conserved. The -35 region is thought to be involved in the initial recognition of the promoter by Rpol (Schaller et al, 1975). DNA protection studies using mild DNAase I digestion of Rpol:promoter complexes (footprinting) indicate that the enzyme covers some 60-70 bp upstream of the mRNA initiation site (Schmitz and Galas, 1979). Further similar studies using chemical and physical DNA modifying agents suggested that Rpol interacts most strongly with the Pribnow box and the -35 region, emphasising their importance in promoter function (Siebenlist et al, 1980). Nucleotide sequences upstream of the -35 region are generally A/T rich, but show no other homologies.

A further sequence homology might be expected between promoters

regulated by cAMP-CAP complex. DNA binding and protection studies on three such promoter regions demonstrate that cAMP-CAP complex binds: 1) *lac* operon DNA at a symmetrical sequence lying -55 to -70 bp upstream of the mRNA start site (Majors, 1976b; Simpson, 1980), 2) *ara* operon DNA at two sites, -78 to -107 bp and -121 to -146 bp upstream of the *araBAD* mRNA start site (Ogden *et al*, 1980), 3) *gal* operon DNA at a site -24 to -50 bp upstream of the P_1 initiated mRNA start site (Taniguchi *et al*, 1979; Galas and Schmitz, 1978). The *gal* promoter possesses an analogous sequence to the *lac* binding site but DNA from this region has been shown to be non-essential for cAMP-CAP dependent *gal*-transcription (Taniguchi *et al*, op.cit.). The *araBAD* promoter has no such homology around the -55 to -70 region. A computer search of sequences around these binding sites reveals a consensus sequence, TGTCACACTTT, which may be important in cAMP-CAP binding. The *ara* and *gal* binding sites conform well to this sequence, but the *lac* binding site is less consistent (Tanaguchi *et al*, op.cit.).

How can we reconcile the different locations of the binding sites observed? Clearly, the means by which cAMP-CAP influences Rpol at these sites may differ. For example, if we assume a cAMP-CAP-Rpol interaction, then in the *gal* operon the binding sites for these proteins are close enough for such contact to occur. On the other hand, in the *ara* operon, the binding sites are well separated and such contact would be unlikely to occur. Ogden *et al* (op.cit.) have proposed a model for *ara* by which cAMP-CAP binding could influence Rpol. This utilises the *araC* protein (inducer form, $paraC^{ind}$) as a "bridging" protein between Rpol initiating transcription at the *araBAD* promoter and cAMP-CAP bound at -78 to -107 bp. They demonstrate the binding of $paraC^{ind}$ to the region between

cAMP-CAP and Rpol protected sequences and suggest the sequential binding of cAMP-CAP, paraC^{ind}, and Rpol to enhance initiation at the araBAD promoter. Events at the lac promoter may be more complicated. The binding site here does not conform well to the consensus sequence and exhibits a dyad symmetry. Contact between cAMP-CAP, bound at -70 to -55, and Rpol has been suggested by Simpson (op.cit.) following his observations that cAMP-CAP protects DNA up to -47 bp and Rpol protects DNA down to -43 bp. However, there is evidence that cAMP-CAP binding may involve another region (-17 to -30) which overlaps a subset of the consensus sequence, ACACTTT (-28 to -34) (Dickson et al, 1977). Co-operative binding of cAMP-CAP to these two regions may compensate for their deviation from the consensus sequence and allow direct contact between cAMP-CAP and Rpol.

Alternatively, cAMP-CAP binding may destabilise the helix allowing Rpol to initiate transcription more easily. This has been called "telestability" and would remove the requirement for protein-protein interactions. McKay and Steitz (1981) proposed that cAMP-CAP may destabilise the promoter region by binding to a left-handed, B-DNA conformation. This predicts a reduction in the winding number of closed circular DNA following cAMP-CAP binding. However, Kolb and Buc (1982) attempted to measure this parameter at cAMP-CAP-bound lac and gal promoters on small covalently closed DNAs, but were unable to observe a reduction in winding. Finally, the reader is referred to the "mini-review" by Adhya and Garges (1982) for a more detailed discussion of how cAMP-CAP may act at these sequences.

Mutant promoter sequences supplement information gained from wild-type sequence comparisons. Most promoter point mutations

affect promoter function adversely (i.e. promoter down mutations) and cluster in either the -10 or -35 region. Of those in the -10 region, almost all occur within the three most highly conserved positions of the heptamer. Mutations which enhance the activity of promoters ("promoter up" mutations) are also concentrated in the -10 or -35 regions. These are generally found to increase homology to the consensus sequence (Rosenberg and Court, 1979).

Termination essentially consists of three steps: the elongation process ceases; the RNA transcript is released; and the Rpol:DNA complex dissociates. The relative order of the last two steps is unclear but with T7 tec1 in vitro, the given order appears to be correct (O'Hare, 1978).

The growing number of terminator sequences characterised allows similar speculation concerning sequences important in the termination event. I have compiled a list of a number of these sequences (Fig. 1.3). Three common features become evident: 1) a region of hyphenated dyad symmetry (i.e. an inverted repeat sequence) generally precedes the termination site, 2) the 3'-terminus of the RNA transcript frequently contains a series of U-residues, 3) the region preceding the termination site is usually, but not invariably, G/C-rich. There is substantial variation among these terminators both in the length of the sequence repeated (i.e. stem length) and distance between the repeats (i.e. loop size). Termination generally occurs 20(±4) nucleotides downstream of the centre of dyad-symmetry, ending rarely at a unique residue, but rather "stuttering" over two or more adjacent nucleotides (Rosenberg and Court, 1979).

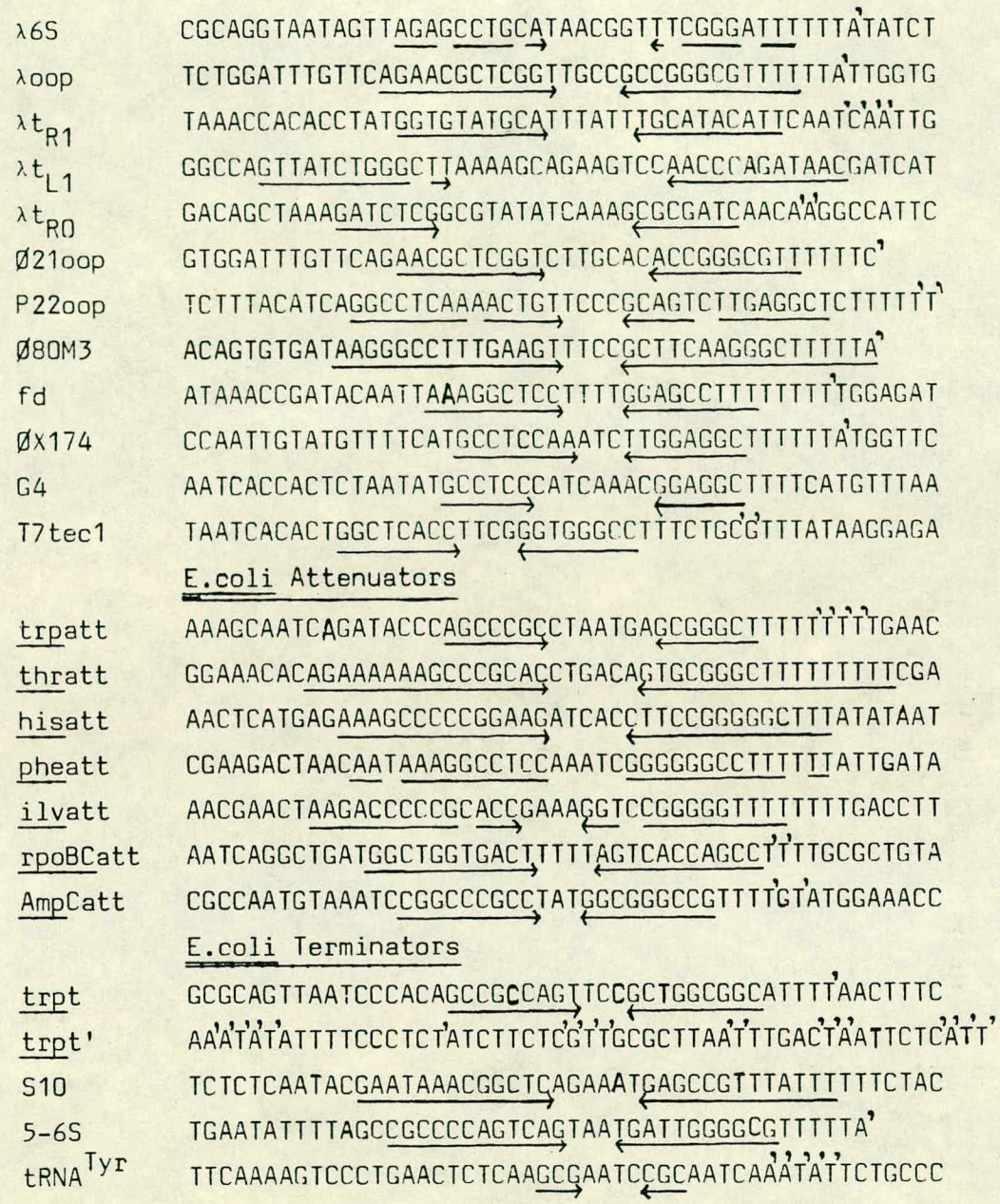
These sites exhibit a wide range of overall efficiency (measured in different systems, and therefore possibly not simply comparable)

Fig. 1.3 Terminator sequences.

The strand with the same nucleotide sequence as the messenger is indicated, 5' to 3', left to right. The region of hyphenated dyad symmetry is underlined where present. The mark (') shown in some sequences indicates site(s) where termination has been shown to occur *in vivo* and/or *in vitro*.

References for these sequences and mRNA endpoints are as follows: λ6S, *in vitro* and *in vivo*; λoop, *in vitro* and *in vivo*; λt_{R1}, *in vitro* and *in vivo*; Ø21 oop, *in vitro*; P22 oop, *in vitro*; Ø80M3, *in vivo*; fd, *in vitro*; ØX174; G4; *trp* att, *in vitro* and *in vivo*; *thr* att, *in vitro*; *his* att, *in vitro*; *phe* att, *in vitro*; S10; 5-6S, *in vivo*; $tRNA^{Tyr}$, *in vitro* (see review by Rosenberg and Court, 1979); λt_{L1} (Drahos and Sybalski, 1981); λt_{RO}, *in vitro* (Calva and Burgess, 1980); T7 tec1, *in vitro* (Dunn and Studier, 1980); *ilvatt* (Lawther and Hatfield, 1980); *rpoBC* att, *in vivo* (Barry *et al*, 1980); *trpt*, *in vitro* and *in vivo*; *trpt'*, *in vitro* and *in vivo* (Wu *et al*, 1981; Platt, 1981); *AmpC* att, *in vitro* (Jaurin *et al*, 1981).

Fig. 1.3 Phage Terminators



E.coli Attenuators

E.coli Terminators

and of dependence on rho factor for termination. The structure of those that function efficiently *in vitro* in the absence of cofactors, and current thoughts on the termination process, led Farnham and Platt (1980) to propose a model for rho-independent termination. They propose that Rpol, in the absence of additional factors, requires both a hairpin structure and a 3'-terminal stretch of uridine residues in the transcript in order to terminate efficiently *in vitro*.

The inverted-repeat sequence could be involved in base pairing in DNA, RNA or DNA:RNA structures. The former and the latter seem unlikely in the light of several observations:

1) If GTP is replaced by ITP in the transcript of the *trp* attenuator, termination does not occur. I/C base pairs are weaker than G/C base pairs. This implicates the involvement of the *transcript* in the termination event (Lee and Yanofsky, 1977).
2) Mutations in the *trp*-attenuator that decrease termination (GC to AT) could destabilise RNA:RNA or RNA:DNA structures (Stauffer *et al*, 1978).
3) Mutations in the *trp* attenuator (GC to CG) and λt_{R1} (AT to GC) exist that decrease the efficiency of termination and should not disrupt RNA:DNA structures (Stauffer *et al*, 1978; Rosenberg *et al*, 1978).

Stem-loop formation in the transcript, therefore, seems to be a requirement for termination. Its contribution to the termination event may be revealed by indications that RNA hairpins elicit pausing by Rpol (Rosenberg *et al*, 1978; Farnham and Platt, 1981). Pausing occurs at the λt_{R1} terminator (Rosenberg *et al*, 1978), at two potential hairpin structures in the *trp*-attenuator (Farnham and Platt, 1981), and at the λt_{R2} terminator in the presence of *nusA* protein

(Greenblatt *et al*, 1981). The importance of the hairpin is emphasised by the large number of mutations affecting termination that are found to affect its sequence (Wulff, 1976; McDermit *et al*, 1976; Rosenberg *et al*, 1978; Stauffer *et al*, 1978; Yanofsky, 1981; Hong *et al*, 1971; Gardner, 1979).

Whether or not RNA-release occurs following pausing may be determined by RNA:DNA interactions immediately following the hairpin. Martin and Tinoco (1980) suggest that a run of uridine residues downstream of the hairpin would facilitate dissociation of the transcript from the template, since a poly rU-dA hybrid of this type is particularly unstable. Should no such run of uridines be present, Rpol might merely resume elongation after a pause at the hairpin structure. Indeed, the deletion of the terminal 4 TA base pairs in the run of 8 present in the *trp* attenuator (*trp* 1419) abolishes termination *in vitro* and reduces it *in vivo* (Bertrand *et al*, 1977). However, termination does occur at this mutant sequence if a mutant (*rpo*203) polymerase is used, which is presumably less demanding than wild-type enzyme regarding the number of uridine residues present (Farnham and Platt, 1980). Furthermore, the use of base analogues such as Br-UTP and allylamine-UTP which form stronger base pairs with A than does UTP, elevates readthrough at the *trp*-attenuator.

The above observations suggest a mechanism for termination in the absence of other cofactors that involves both RNA:RNA and RNA:DNA pairing. Termination of transcription would depend positively on the rate of formation or stability of the RNA hairpin, and negatively on the stability of the distal rU:dA hybrid region. These ideas have been further supported by the study of synthetic, functional, termination sites possessing the two features discussed here (Platt, 1981).

Examination of terminator sequences reveals a class that do not possess a series of uridine residues at the 3' end of the transcript. These terminators (λt_{R1}, trpt', $tRNA^{Tyr}$ and λt_{RO}) all exhibit an absolute dependence on rho factor for termination (Rosenberg et al, 1978; Platt, 1981; Küpper et al, 1978; Calva and Burgess, 1980). This justifies their classification as fully rho-dependent terminators, the distinction between rho dependence and independence being a hazy one. [For example, rho does not affect the cessation of RNA elongation at the trp attenuator in vitro, but may stimulate the release of the leader transcript (Fuller and Platt, 1978). It may, therefore, merely enhance termination without increasing it.]

The structure of rho-dependent terminators is interesting, but somewhat confusing. A comparison of the λt_{R1}, $tRNA^{Tyr}$ and λt_{RO} terminators at the sequence level reveals: 1) a weak dyad symmetry preceding the termination point, 2) an A/T-rich sequence spanning the RNA terminus, 3) the common sequence CAATCAA in λt_{R1} and $tRNA^{Tyr}$ with the subset ATCAA in λt_{RO} just preceding the stop site. Rho may act to confer instability at this site. Indeed, an unusual feature of the $tRNA^{Tyr}$ terminator is the triple repeat of a near identical 178 bp sequence (Küpper et al, 1978). Termination in vitro occurs not in the first repeat but in the second and could potentially occur in the third. The base differences between these repeats may reveal important sequences for rho-dependent termination. The most likely important difference lies within the common sequence ATCAA mentioned earlier. The second and third repeats conform to this sequence whereas the first exhibits the sequence ATTAA. This C/T base change is the closest to the termination region and supports the suggestion that this sequence is important in rho-dependent termination.

The trpt' sequence has none of these features, but is A/T-rich and causes heterologous termination over a stretch of 50 nucleotides in the presence of rho. Inclusion of nusA protein in this transcription reaction yields only one major RNA product suggesting a number of cofactors maybe required for termination here. The lack of similarity between trpt' and other transcription terminators may also suggest a distinct termination event at this site. Indeed, two functional termination sites exist at the end of the trp operon: 1) a weak rho-independent terminator with a dyad-symmetry, followed by a series of uridine residues, (trpt); 2) the above strong, rho-dependent terminator, 280 nucleotides downstream, (trpt'). The 3' terminus of mRNA isolated in vivo exhibits a sequence corresponding to trpt (Wu and Platt, 1978). This could be the result of a rho-dependent interaction between the two terminators in vivo or termination at both sites followed by a processing of the longer transcript. If we can extend observations in vitro to the situation in vivo, then the former possibility seems unlikely. Wu et al (1981) have shown that when cloned individually next to the trp promoter, in vitro transcripts do not differ from those obtained with the terminators in tandem configuration. This would rule out an interaction in vitro and presumably in vivo. Hence, the explanation of their behaviour and importance requires further study.

Concluding this section, it should be emphasised that it is not, as yet, possible to recognise promoters or terminators merely from nucleotide sequence. Their existence and an estimation of their efficiencies requires the use of gene fusions utilising e.g. gal, lac or trp genes to monitor their activities. Experimental results have implicated RNA:RNA and RNA:DNA interactions in the termination

event. Despite this, several questions remain to be answered: 1) How does Rpol recognise the RNA hairpin presumably behind it? 2) How does the hairpin promote pausing? 3) Is this pausing involved in a mechanism for regulating the elongation rate of Rpol? 4) Do ribosomes have an influence on this and/or other steps of termination? 5) What is the significance of the multiple terminators at the ends of the trp and $tRNA^{Tyr}$ operons? 6) How do rho-dependent and rho-independent termination differ? 7) How does commitment to RNA release arise, following pausing? 8) Do other cofactors influence these events besides those discussed?

The probing of molecular interactions between proteins involved in recognising transcription terminators and the sequences they recognise may provide some answers to the above questions. Meanwhile, our ignorance is still great.

TERMINATION AS A CONTROL MECHANISM FOR GENE EXPRESSION

Efficient transcriptional termination is essential at the ends of operons to prevent transcriptional interference affecting expression of genetic information downstream. The terminator's role in the transcription cycle has in the past, as its name implies, been confined purely to this. The discovery of terminators proximal to, within, and between coding sequences has led to a reappraisal of termination as an important means of controlling gene expression in E.coli. This section will discuss three instances of terminator-modulated transcription:

Attenuation

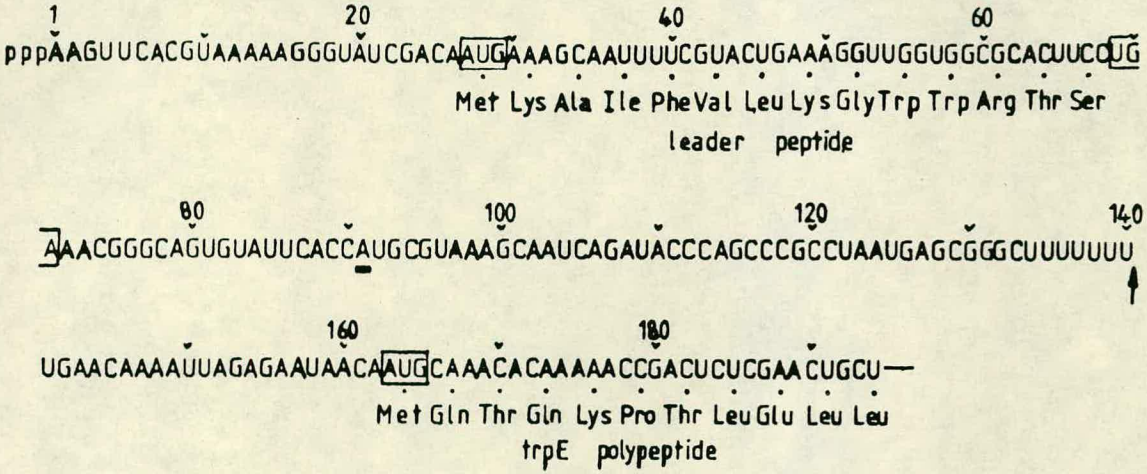
Early studies on the control of expression of the trp operon concentrated on the effects of the 47 kdal homodimer product of the unlinked trpR gene, the aporepressor (Gunsalus et al, 1979). When

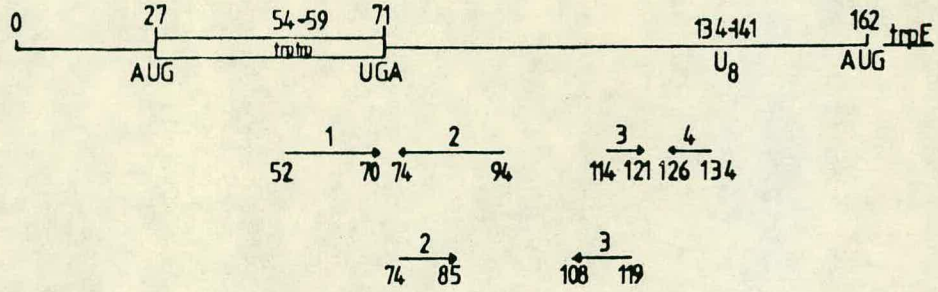
complexed with tryptophan, this prevents initiation at the trp promoter by binding to the overlapping operator site. But several early observations suggested that this was not the sole means of trp regulation: 1) tryptophanyl-tRNA synthetase mutants display regulatory anomalies; addition of tryptophan to these causes only partial repression of trp operon expression (Hiraga et al, 1967; Kano et al, 1968; Ito, 1972). 2) the addition of tryptophan to trp-starved cells inhibits transcriptional initiation and elongation in the early region of the operon (Imamoto, 1968). 3) mutants lacking a functional repressor are still able to exhibit some degree of regulation over the trp-operon (Baker and Yanofsky, 1972; Morse and Morse, 1976). 4) certain trp-deletions, contained wholly within the transcribed region, display an unexpected six-fold stimulation of the distal genes of the operon (Jackson and Yanofsky, 1973; Bertrand et al, 1976). 5) the first 140 nucleotides of the trp mRNA are more abundant than more distal segments, presumably due to "premature" termination of some transcripts (Bronson et al, 1973; Bertrand et al, 1976; Squires et al, 1976). These observations imply a complex mechanism of control for trp expression involving tryptophan, aporepressor, trp-tRNA synthetase or its product, and a site of transcriptional discontinuity in the early region of the operon.

The nucleotide sequence of the region proximal to the first structural gene (trpE) has several interesting features (Squires et al, 1976; Lee et al, 1978) (see Fig. 1.4). The 162 nucleotides lying between the 5'-end of the trp mRNA and the AUG of trpE are known as the leader transcript. It potentially encodes a 14 amino-acid peptide, with tandem tryptophan residues at positions 10 and 11, and a UGA stop codon at +69 to +71. Between here and the end

Fig. 1.4

The nucleotide sequence of the 5'-end of *trp* messenger RNA and the proposed secondary structures in the *trp* leader transcript. The sequence shows: the boxed AUGs where translation starts and the boxed UGA where it stops; the predicted amino acid sequence of the *trp* leader peptide; the 3'-terminus of the paused transcript at nucleotide 90 (underlined); the 3'-terminus of the *trp* leader transcript at nucleotide 140-141 (arrowed). Beneath this, the line diagram represents the *trp* leader transcript with the coordinates of regions of interest given as nucleotides from the mRNA start point. Inverted repeat sequences are represented by arrows and referred to as 1, 2, 3 and 4 for simplicity. To the right are shown the proposed secondary structures in the *trp* leader transcript together with their calculated free energies of formation (in Kcal mol^{-1}) (from Oxender *et al*, 1979).


1
20
40
60
pppAAGUUCACGUAAAAAGGGUAUCGACAAUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUG
Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser
leader peptide
80
100
120
140
AAACGGGCAGUGUAUUCACCAUGCGUAAAGCAAUCAGAUACCCAGCCCGCCUAAUGAGCGGGCUUUUUUU
160
180
UGAACAAAAUUAGAGAAUAACAAUGCAAACACAAAAACCGACUCUCGAACUGCU—
Met Gln Thr Gln Lys Pro Thr Leu Glu Leu Leu
trpE polypeptide


0
27
54-59
71
134-141
162
trpE
trpLw
AUG
UGA
U8
AUG
1
2
3
4
52
70
74
94
114
121
126
134
2
3
74
85
108
119

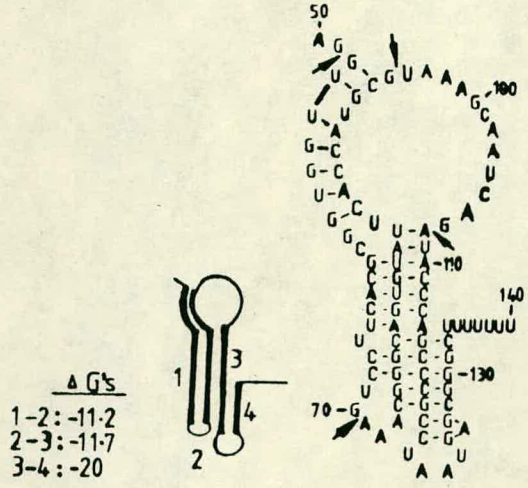

ΔG's
1-2: -11·2
2-3: -11·7
3-4: -20
1
2
3
4
50
100
110
130
140
70
UUUUUUU

Fig 1·4

of the leader is a GC-rich region of dyad-symmetry (3:4, +114 to +121 and +126 to +134) followed by a run of 8 TA base pairs (+134 to +141), similar to the rho-independent transcription terminators discussed earlier. Ribosomes bind tightly around the AUG of the potential 14 amino-acid peptide (Platt et al, 1976) and although leader peptide synthesis in vivo has not been demonstrated, studies of fusions of the initial section of the trp leader directly to trpE have shown that translation can begin at this point in vivo (Miozzara and Yanofsky, 1978). Furthermore, the possible terminator functions in vitro and in vivo in the absence of rho. Transcription in vitro yields products ending with 7 or 8 U-residues at positions +140 and +141 (Lee and Yanofsky, 1977). Similarly, transcription in vivo produces transcripts ending in 4 to 8 U-residues, due to stopping between +137 and +141. At early times of transcription in vitro, some RNA products end at position +90, distal to a second possible dyad-symmetry (1:2, +52 to +70 and +74 to +94). Kinetic studies indicate that this is not the result of a true termination event, but is caused by Rpol pausing at this point. The paused transcript resumes elongation to become the terminated leader transcript or the readthrough transcript (Yanofsky, 1981). A third, dyad-symmetry, is observed between the first two symmetries (2:3, +74 to +85 and +108 to +119). Formation of this hairpin would exclude the formation of both hairpins described earlier. Termination does not occur after this symmetry. Finally, a fourth possible RNA hairpin structure exists involving an early segment, containing the leader ribosome binding site, and a distal segment of the transcript. If this pairing were to occur it would exclude ribosome-binding and hence translation of the leader (Yanofsky, 1981). Free energy calculations suggest that all of the possible hairpins

are plausible. Moreover, assessment of leader RNA structure by hydrolysis with the structure-sensitive T_1-RNAase indicate that at least the first three possible structures are involved in base pairing (Lee and Yanofsky, 1977; Oxender et al, 1979). Furthermore, the importance of the 3:4 hairpin is emphasised by the properties of a number of deletions: one, whose promoter-proximal end-point is in the centre of the potential stem-loop, completely abolishes termination; one, which shortens the run of TA base pairs from 8 to 4, reduces termination efficiency (Δtrp1419); and one, whose nearer end-point is 11 nucleotides beyond the normal RNA stop-site, has no detectable effect on termination (Bertrand et al, 1977).

Efficient termination is observed in vitro, when translation is absent, or upon starvation for amino acids located early in the leader peptide. Similarly, a mutation (trpL29) which changes the leader region AUG initiated codon to AUA, reducing ribosome initiation, leads to increased termination in vivo as well as a reduced response to starvation (Zurawski et al, 1978). Furthermore, mutations in trpS, leading to defective charging of $tRNA^{Trp}$, and tryptophan (or arginine) starvation lead to relief of termination at this site (Morse and Morse, 1976; Bertrand and Yanofsky, 1976; Zurawski et al, 1978), This suggests that charging of $tRNA^{Trp}$ is involved in trp operon expression although other mutants exist which suggest that defective tRNA-ribosome interactions are influential in this (Yanofsky and Soll, 1977). Nevertheless, these findings imply strongly that translational effects (clearly involving the leader peptide) are important in trp operon expression.

Current thoughts on the events regulating trp operon expression are based on a model proposed by Oxender et al (1979). The mechanism has been designated "attenuation" and occurs when an Rpol molecule

escapes repressor control at the trp promoter. Transcription of the leader region is accompanied by ribosome attachment at the leader polypeptide ribosome binding site. Transcription and translation continue until Rpol reaches the first dyad-symmetry, pausing at position +90. This may allow the following ribosome to approach the polymerase molecule and so facilitate the synchronisation of ribosome with Rpol movement. When transcription resumes, the translating ribosome may stall at one of the adjacent tryptophan codons (if the cell is deficient in tryptophan), or move to the translation stop codon (if tryptophan is plentiful). The position of the ribosome on the transcript determines which alternative RNA secondary structure forms, and hence the decision to terminate or continue transcription. Stalling of the ribosome on the trp-codons allows stem-loop 2:3 to form, excluding stem-loop 3:4 formation, and allows Rpol to extend transcription into the trp structural genes. Termination of translation at the translation stop codon prevents stem-loop 2:3 formation, thus allowing stem-loop 3:4 formation and so promoting transcriptional termination. Total absence of translation has a similar effect, allowing formation of stem-loops 1:2, and 3:4, the latter being essential for transcription termination. Whether termination occurs or not, the leader ribosome-binding site may be blocked by pairing with the distal transcript segment, which would prevent further (redundant) leader translation.

This model explains the observation that starvation for arginine is also able to relieve trp attenuation in vivo (Zurawski et al, 1978), since stalling at the arg codon distal to the trp codons should allow stem-loop 2:3 formation as before. Similarly, mutations that relieve or enhance attenuation disrupt or promote

(respectively) the formation of stem-loop 3:4 (for review see Yanofsky, 1981). The exploitation of transcriptional termination and transcription-translation coupling provides an elegant control system for trp-operon expression. The effect of repression exerts the greatest quantitative effect in vivo, reducing transcription by up to 70-fold, and responding to tryptophan levels exclusively. Attenuation, on the other hand, reduces transcription in vivo by up to 10-fold and is influenced by the activity of tryptophanyl-tRNA synthetase, the amount of $tRNA^{Trp}$, and the overall rate of protein synthesis. Tryptophan starvation affects both repression and attenuation and can, therefore, reduce transcription of the trp-operon by up to 700-fold. The combined action of the two mechanisms allows E.coli to respond to both external and internal constraints relevant to expression of the trp-operon. Finally, several other amino-acid biosynthesis pathways in gram-negative bacteria utilise very similar mechanisms to regulate expression of relevant enzymes (e.g. his; thr; ile; phe) (Yanofsky, 1981).

Antitermination

Transcriptional termination has a major influence on the regulation of expression of the bacteriophage λ genome (Hershey, 1971; Roberts, 1976; Das et al, 1976). Following phage infection, the host Rpol transcribes the "early" genes of λ initiating transcription of two major promoters, P_L and P_R, within the immunity region (see Fig. 1.1). Leftward and rightward transcription is limited by rho-dependent transcription terminators, t_{L1} and t_{R1}, flanking the immunity region. N is the first gene of the major leftward operon (transcribed from P_L) and its product, N protein, is required to antagonise termination at t_{L1} and t_{R1} and, hence, to permit efficient transcription of distal genes (Taylor

et al, 1967; Salstrom and Sybalski, 1978a).

Readthrough is also stimulated at the λt_{L2} termination site, which is apparently rho-independent, thus pN is truly an anti-termination factor and not merely an anti-rho factor (Gottesman et al, 1978). This antitermination is phage specific (the N-gene product of the lambda-like phage 21 will not complement λN-mutants), suggesting a specific interaction with site(s) on the phage genome (Friedman et al, 1973). Deletions have defined such sites to the left of P_L (nut L, for N-utilisation) and to the right of the cro gene (nut R) (Adhya et al, 1974; Salstrom and Szybalski, 1976). Furthermore, the nucleotide sequences of these regions reveal two sites of near identical two-fold rotational symmetry (16 of 17 bp) appropriately positioned to be nutL and nutR. The occurrence of nutL mutations within this 17 bp sequence suggests that these two sequences are indeed involved in N protein activity (Rosenberg et al, 1978; Salstrom and Szybalski, 1978b) . Interestingly, these sites exhibit a potential for hairpin formation and possess distal sequences similar to the CAATCAA noted in some rho-dependent terminators. Transcription termination does not appear to occur here, but the stem-loop might be involved in Rpol pausing to facilitate pN-Rpol interaction (assuming that pausing is a general effect of hairpin structures). The antitermination effect of pN can be observed at nearly all terminators downstream of a nut site. This has been demonstrated, for example, at trpt in λ/trp fusions, and by suppression of a variety of polar mutations (Segawa and Imamoto, 1976; Adhya et al, 1974).

Evidence for an N protein-Rpol interaction comes from the observations of Greenblatt and suggests a mechanism by which this may occur. The interactions he has shown between nusA protein,

Rpol, and N protein (discussed earlier) suggest that nusA protein may mediate binding of N protein to Rpol, thereby creating the transcriptional juggernaut proposed by Adhya (1974). This suggestion is supported by the demonstration that the products of the nusA and nusB genes are required for phage N function to act in vivo (Ward and Gottesman, 1981). Should this be the case, we can speculate that pnusA may act in a similar way to bind other cofactors to Rpol, e.g. rho (as discussed earlier) and other undiscovered cofactors.

N protein also stimulates transcription from P_L and P_R 2-4 fold (Hu et al, 1979; Hu and Szybalski, 1979) and from other promoters derived from both λ and pBR322 (Drahos and Szybalski, 1981). This has yet to be demonstrated to be a general effect of N protein and its mechanism is obscure.

Drahos and Szybalski (1981) have investigated the nucleotide sequences involved in pN action by studying each module separately and in various combinations. They have utilised the galK-fusion system for monitoring transcription signals developed by McKenney et al (to be described later in this thesis) to examine the relationship between their structure and function. They demonstrate: 1) the termination activity of λt_{L1}, galactokinase activity being reduced by 80%; 2) N protein stimulation of transcription from P_P (pBR322) and P_R' (λ); 3) a 25 bp sequence carrying the 17 bp dyad-symmetry termed nutL restores 60% of pN-stimulated transcription if inserted in the correct orientation. Such antitermination will only occur at terminators downstream of a correctly orientated nutL sequence and no effect on transcription is observed in the absence of pN-function. The overall results strongly support the proposal that the 17 bp dyad-symmetry constitutes a site required for pN

activity and clearly demonstrate the interactions between various components of the λ antitermination system.

Nevertheless, the mechanism of pN-mediated antitermination remains open to speculation. Adhya and Gottesman (1978a) proposed a number of mechanisms for pN action, variations of these are: 1) pN interacts with Rpol (possibly via pnusA and pnusB) producing a conformational change in the polymerase such as to disfavour termination. 2) pN may interact with Rpol as above, but mediate antitermination by antagonising ribosome discharge at stop codons in such a way that rho can no longer gain access to Rpol (via nascent RNA) to catalyse termination. The ability of pN to antagonise termination even at rho-independent sites (e.g. λt_{L2}) would require further explanation. We can speculate, for example, that undischarged ribosomes might prevent formation or function of RNA stem-loops required for termination. 3) pN may affect the structure of nascent RNA directly so as to prevent termination. Evidence supporting the involvement of ribosomes in antitermination comes from the observation that some ribosomal protein mutants affect N protein function (Friedman et al, 1981). Such mutations lie in the nusE gene which codes for the ribosomal protein S10. Friedman et al have suggested that pN prevents transcription termination by coupling Rpol to the ribosome. Alternatively they suggest S10 may have an indirect role in antitermination by regulating the synthesis of a protein that is directly involved e.g. pnusA, pnusB or pN. However, the observation that N protein appears to function in the absence of ribosomes in vitro argues against their involvement in the antitermination event (Ishii et al, 1980).

The available evidence does not exclude any of these suggestions. The affinities of Rpol, pnusA, and pN for each other do, however,

suggest the formation of a tripartite complex and a continuous association during subsequent transcriptional activities. At all events, the balance between termination and antitermination by pN plays a subtle and evidently important role in the regulation of λ gene expression.

Polarity

The final means by which termination affects gene expression to be discussed is the phenomenon of polarity. Some mutations in structural genes affect not only the product of the mutated gene, but also reduce or abolish the expression of other, distal genes within the same operon (Jacob and Monod, 1962). Such mutations are of two types: 1) the generation of an in-phase nonsense codon by point mutation, insertion, or deletion of DNA (nonsense polarity); 2) the insertion/deletion of DNA creating a new translation frame with a stop codon either within an insertion, or distal to the site of mutation (frame-shift polarity). Translation will cease at the newly created site, in the absence of nonsense suppressors, and RNA sequences distal to it will not be translated, unless reinitiation of translation occurs (Platt et al, 1972).

Adhya and Gottesman (1978b) proposed a model to account for polarity based on two major features: there are cryptic rho-dependent transcription terminators within operons; and these sites are immune to rho-activity whilst the nascent mRNA is being translated. The model requires a tight coupling of transcription and translation and is supported by the existence of secondary mutations that relieve polarity without restoring translation of the mutant gene (Morse and Primakoff, 1970). Several such mutations have been shown to affect the termination factor rho, suggesting that rho-dependent transcriptional termination is involved in polarity

(Richardson *et al*, 1975 ; Korn and Yanofsky, 1976; Ratner, 1976b). Polarity relief by many different rho mutants also suggests that mRNA degradation does not play a major role as this should be unaffected by defective rho. Furthermore, the demonstration of rho-dependent terminators within the *trp*, *lac* and *gal* operons and the λ*cro* gene provides evidence for the existence of termination sites required by the model. Similarly, the relief of polarity by p*N*-mediated antitermination supports the suggestion that polarity is primarily due to premature termination of transcription (Adhya *et al*, 1974; Franklin, 1974).

Their model proposes that when an operon is being expressed, the nascent mRNA is normally covered with ribosomes. Potential rho-dependent transcription termination sites may be present within the nascent mRNA, but are prevented from functioning because rho cannot gain access to the RNA. Creation of a nonsense codon within the transcript leads to loss of ribosomes and exposure of distal mRNA. Rho attaches to the latter transcript, possibly at preferred sites, and by "diffusing" along the chain,it contacts an Rpol molecule paused at the next potential rho-dependent terminator. Rho subsequently induces termination here, leading to release of the mRNA chain and dissociation of the Rpol molecule. In the light of present knowledge we might extend this model to permit rho-*independent* termination sites a potential role in polarity: ribosomes may prevent stem-loop formation in the nascent mRNA and so oppose rho-independent termination and possibly even pausing.

Polarity can prevent the wasteful synthesis of large regions of untranslatable mRNA, or the expression of distal gene products which are often useless in the absence of the mutated gene product. Thus, there may be a selective pressure for the retention of cryptic transcription terminators within genes.

A SUMMARY OF THE TRANSCRIPTIONAL BEHAVIOUR OF COLIPHAGE T7

The original aim of these studies required the cloning of a fully rho-independent terminator for E.coli RNA polymerase, from the genome of bacteriophage T7. This parasite of E.coli has been well studied genetically and biochemically (for reviews see Studier, 1972; Hausmann, 1976; Studier et al, 1979; Krüger and Schröder. 1981). In the following section, I shall summarise what is known about the transcriptional properties of coliphage T7.

Coliphage T7

T7 has a genome of linear, double-stranded DNA which is not circularly permuted (Studier, 1972) but has a 160 bp direct-repeat at each end (Dunn and Studier, 1981). The genome comprises some 40,000 bp (McDonell et al, 1977; Dunn and Studier, 1981) and codes for about 30 identified gene products, 20 of which are known to be essential for laboratory growth on a wild-type host (Studier, 1972). Some 15 additional, putative genes have been identified solely by nucleotide sequencing of the genome (Boothroyd and Hayward, 1979; Dunn and Studier, 1981; Dunn and Studier, 1982). Identified genes of T7 are conventionally numbered according to their relative positions on the T7 genetic map, those more recently discovered being given decimal numbers to indicate their relative positions. Co-ordinates of regions of interest may be expressed in base pairs or, more usually, "T7 units", each of which is 400 bp (1% of genome).

The genome is genetically segregated into an early region and a late region. The early region is defined by three strong promoters for E.coli RNA polymerase between 1.25% and 1.88% and a transcriptional termination site for E.coli RNA polymerase at 18.89%. Nine genes, numbered from 0.3 to 1.3, have been mapped within this

Fig. 1.5 Simplified genetic map of the T7 genome.

The major coding sequences in the "early" region are indicated by boxes, with the relevant gene number above each box. Transcription and protein products are indicated below the line. The transcriptional signals defining the "early" region (promoters A_1, A_2 and A_3; terminator tec1) are indicated above the line. The open arrow indicates the direction of transcription from A_1, A_2 and A_3. The "late" region is also boxed. The collective function of each group of genes is recorded within the relevant box.

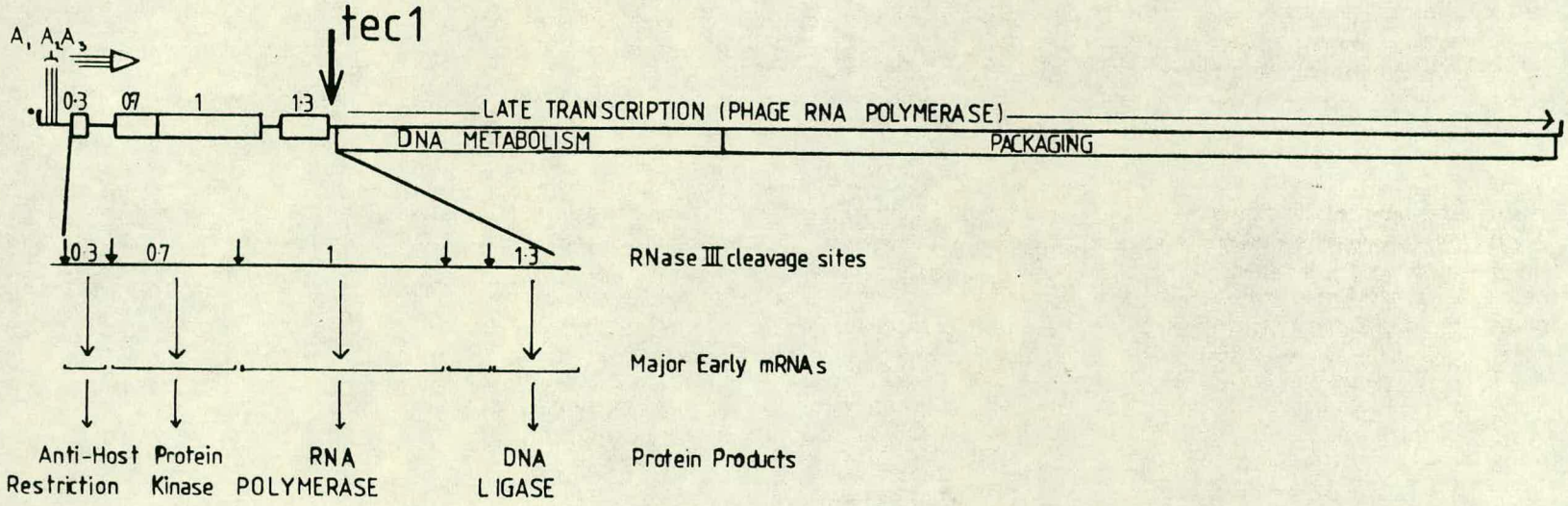

A1 A2 A3
tec1
0·3
0·7
1
1·3
LATE TRANSCRIPTION (PHAGE RNA POLYMERASE)
DNA METABOLISM
PACKAGING
0·3
0·7
1
1·3
RNase III cleavage sites
Major Early mRNAs
Anti-Host Restriction
Protein Kinase
RNA POLYMERASE
DNA LIGASE
Protein Products

Fig 1·5

region and are involved in early stages of infection (Studier et al, 1979). The late region overlaps the early region somewhat and is transcribed by a T7 encoded RNA polymerase. It can be further sub-divided on the basis of the time-course of gene expression into the Class II genes, which generally specify proteins involved in DNA metabolism, and the Class III genes, which generally specify proteins involved in DNA maturation, and structural proteins of the phage particle (Studier, 1972).

The nucleotide sequence of the phage genome has revealed a number of interesting features including: the structure of the three promoters for E.coli Rpol at the start of the early region (Siebenlist, 1979); the structure of the E.coli Rpol-terminator at the end of the early region (Dunn and Studier, 1980); the structure of ribonuclease III targets within the early region (Robertson et al, 1977; Rosenberg and Kramer, 1977; McConnell, 1979; Oakley and Coleman, 1977; Dunn and Studier, 1981); the structure of a number of promoters for the phage specified T7 Rpol (Rosa, 1979; Oakley and Coleman, 1977; Boothroyd and Hayward, 1979; Panayotatos and Wells, 1979; Dunn and Studier, 1981); and the structure of the primary origin of DNA replication (Saito et al, 1980). The lack of any real sequence homology between promoters for E.coli Rpol and T7 Rpol fits the observation that each polymerase recognises and interacts only with its own promoters. Despite the similar activities of the two proteins, the only similarity between the sites they recognise is a general A/T-richness upstream of the mRNA start site. Sequence information also reveals that T7 DNA is remarkably efficient as a carrier of coding information for proteins. The coding sequences are interrupted by only small stretches of untranslated nucleotides, many of which contain some other type of

genetic signal. It is noteworthy that these regions include promoters for both polymerases, the early region terminator and the origin of replication. Perhaps translation of RNA from these regions would interfere with their activities. Where no such interruptions occur, junctions between genes are typically close; often the termination codon for one protein overlaps the initiation codon for the following protein or the coding sequences actually overlap by several nucleotides. However, extensive overlapping of genes in different reading frames has not been observed in T7.

Class I gene expression

These genes are located within the "early region", and are transcribed from about 0 to 8 minutes after infection at 30°C (McAllister and Wu, 1978). Their products are collectively involved in neutralising the host's defence system and in directing the cell's activities towards the expression of T7 genes.

Purified E.coli Rpol initiates transcription of T7 DNA preferentially at three closely associated promoters, designated A_1, A_2 and A_3, at map positions 1.25%, 1.57% and 1.88% (Stahl and Chamberlin, 1977; Siebenlist, 1979; Dunn and Studier, 1981). With high Rpol:T7 DNA ratios, polymerase molecules will also bind to four other sites on the genome (Stahl and Chamberlin, 1977; Koller et al, 1978; Delius et al, 1973). Designated minor promoters A_0 or D (0.6%), B (3.7%), C (7.7%) and E (92.1%), transcription can be initiated at these sites under certain conditions in vitro (Stahl and Chamberlin, 1977; McConnell, 1979a). The physiological importance of these sites has yet to be established.

Transcription proceeds from left to right using only the r-strand as template (except for A_0) and terminates with up to 95% efficiency at the 18.89% terminator, designated tec1 (Millette

et al, 1970; Peters and Hayward, 1974; O'Hare, 1978). Termination is rho-independent and $^{2}/_{3}$ of the transcripts end with a C-residue, the remaining $^{1}/_{3}$ terminating at the subsequent G-residue (Dunn and Studier, 1980). McConnell has shown that, in the event of initiation at the C promoter in vitro, rho-independent termination can occur at two or more sites before the start of the gene 1 coding region (1979b). The resulting major polycistronic transcript can be cleaved by purified ribonuclease III, mimicking the in vivo situation, to yield five independent mRNAs and at least four short untranslated species (Rosenberg et al, 1974; Robertson et al. 1977). Both the independent mRNAs and the polycistronic transcript can be translated, with much the same efficiency, to yield T7 early proteins. The one exception to this is that translation of 0.3-coding region is much reduced in the unprocessed transcript. This could reflect secondary structural effects on initiation of translation due to the extra RNA in the message (Steitz and Bryan, 1977).

Transcription in vivo is apparently very similar to that in vitro. Initiation occurs almost exclusively at the A promoters (Studier, 1972), although initiation within the "early region" has been reported (Minkley and Pribnow, 1973; Pfennig-Yeh et al, 1978), and proceeds from left to right. In a wild-type host, 50-75% of transcripts will terminate at a site between the 0.4 and 0.7 genes yielding excess of 0.3/0.4 gene mRNA after infection (Hercules et al, 1976; Pfennig-Yeh et al, 1978). Of the transcripts that extend beyond here, a further 50% terminate at a site between genes 0.7 and 1 (Hercules et al, 1976). The remaining transcription complexes proceed to the end of the operon and terminate at tec1, after gene 1.3 (Studier, 1972). The efficiency of termination at this site

is at least 75% (McAllister and Barrett, 1977). Rho mutations do not affect this (Kiefer et al, 1977).

The only class I gene products essential for growth on a wild-type host are those of genes 0.3 and 1. The latter is a single polypeptide, T7-specific-Rpol which is responsible for transcription of the class II and class III genes. The 0.3 gene codes for a protein (MW 14,000) involved in overcoming the host's restriction system (Studier, 1975). The 0.7 gene codes for a protein kinase (MW @ 36,000) which apparently phosphorylates the host Rpol (Zillig et al, 1975). Phenotypically, this gene is required for rapid shut-off of early gene expression (McAllister and Barrett, 1977b). However, shut-off does occur after some delay in T7 0.7^- mutants, evidently as a result of gene 2 product activity (Hesselbach and Nakada, 1977). The mechanism of shut-off remains obscure but is presumably a phosphorylation of the E.coli Rpol and/or direct binding and inhibition of Rpol by the gene 2 product (Hesselbach and Nakada, 1977).

The 1.3 gene codes for a non-essential DNA ligase which is probably involved in DNA synthesis and metabolism (Studier, 1972); T7 1.3^- strains are effectively complemented by the host. Not surprisingly, in view of its function it is expressed in class II as well as class I (Oakley and Coleman, 1977).

Class II and Class III gene expression

The class II gene products are responsible for T7 DNA metabolism and host DNA degradation (Studier, 1972). Transcription of these genes occurs from about 4 to 12 minutes after infection at 30°C (McAllister and Wu, 1978). The class III genes code for structural, assembly and DNA-maturation functions (Studier, 1972). These genes are transcribed from about 8 minutes after infection until the time

of lysis. The delay in their initiation is part of a general pattern almost certainly arising from a phased injection of the phage DNA. The known promoters for T7 Rpol can be divided into two classes, which correspond to the two classes of T7 late proteins. The nucleotide sequences around the RNA start sites for five class II and five class III proteins have been determined (Oakley and Coleman, 1977; Rosa, 1979; Panayotatos and Wells, 1979; Boothroyd and Hayward, 1979). The five class III promoters exhibit a 23 base pair continuous homology that includes the RNA start site, plus other strong homologies. The sequences of the five class II promoters are highly homologous with the 23-base pair class III homology, but each diverges from it in at least 2 and as many as 5 positions. The switch from class II to class III gene expression is not understood but is presumably influenced by these sequence differences and is apparently dependent on the 3.5 gene product (McAllister and Wu, 1978), N-acetylmuramyl-L-alanine amidase known as T7 lysozyme (Inouye et al, 1973). This product is not, however, required for ultimate lysis of the infected cell.

Transcription of both classes of genes initiates at several promoters, that of class II terminating mostly at a site near position 61% whilst that of class III proceeds to the right-hand end of the genome. The distribution of promoters is such that the region from gene 8 through gene 10 is expressed from both class II and class III promoters.

The Present Study

I will describe the cloning of a number of T7 fragments thought to contain transcription terminators including one which I have discovered (in leftward orientation) in the late region. Fusion of

these fragments *in vitro* to *galK* has permitted the measurement of their efficiencies in a defined system *in vivo*, for comparison with earlier estimates. The sequence of the new transcription terminator is also presented.

These constructions should facilitate the future isolation and study of mutant terminator sequences generated *in vitro* in an attempt to clarify the termination event further.

CHAPTER 2 - MATERIALS AND METHODS

1. Growth Media

L-broth contained, per litre: Difco Bacto Tryptone, 10g; Bacto Yeast Extract, 5g; NaCl, 5g; adjusted to pH 7.2.

L-agar contained, in addition: NaCl, 5g; Difco agar, 15 g.

BBL-agar contained, per litre: Baltimore Biological Laboratories trypticase, 10g; NaCl, 5g; Difco agar, 10 g.

BBL top-layer is BBL-agar with only 6.5 g Difco agar per litre.

Bacterial buffer contained, per litre: KH_2PO_4, 3 g; Na_2HPO_4, 7 g; NaCl, 4 g; $MgSO_4.7\ H_2O$, 0.2 g.

Phage buffer contained, per litre: KH_2PO_4, 3 g; Na_2HPO_4, 7 g; NaCl, 5 g; 0.1 M $MgSO_4$, 10 ml; 0.01 M $CaCl_2$, 10 ml; 1% (w/v) gelatin, 1 ml.

Spizizen minimal medium contained, per litre: $(NH_4)_2SO_4$, 2 g; K_2HPO_4, 14 g; KH_2PO_4, 6 g; Trisodium Citrate Dihydrate, 1 g; $MgSO_4.7\ H_2O$, 0.2 g; Final concentrations of supplements were: Sugars or glycerol, 0.2% (w/v); amino acids, 20 µg/ml; vitamin B_1, 2 µg/ml; acid casein hydrolysate, 0.1% (w/v). Spizizen minimal agar contain, in addition: Difco agar, 15 g.

MacConkey agar contained, per litre: Bacto-Peptone, 17 g; Difco Protease Peptone, 3 g; Bacto Bile Salts No. 3, 1.5 g; NaCl, 5 g; Bacto-Agar, 13.5 g; Bacto Neutral Red, 0.03 g; Bacto Crystal Violet, 0.001 g; Carbohydrate source, 10 g.

Galactose-Tetrazolium indicator plates contained, per litre: as for L-agar and in addition: Galactose, 0.8%; Tetrazolium, 0.02 g.

Antibiotics were used at the following concentrations:

Ampicillin	50 µg/ml
Tetracycline	25 µg/ml
Chloramphenicol base	100 µg/ml

TABLE 2.1 Bacterial Strains

Strain	Genotype	Donor
C600	*lac leu thi thr tonA supE*	N.E. Murray
AA125	Δ(*gal-uvrB*) Δ(*lacZ*)M15 Δ(*trpED*) 102 *lacI3 his tonA tsx*	N.E. Murray
MC1000	Δ(*araABOIC-leu*)7697 Δ(*lacIZY*) X74 *araD139 galU galK strA*	M. Casadaban
ED8654	*metB galK2 galT1 lacY* $hsdM^+$ $hsdR^-$ *supE supF*	N.E. Murray
Escherichia coli B	Prototroph	F.W. Studier
N100	*galK2 recA13*	K. McKenney
JM101	Δ(*lac pro*) *thi strA endA sbcB115 supE*/F'*traD36* $proAB^+$ $lacI^+$ *lac*ΔM15	K. Murray

TABLE 2.2 Phage Strains

Strain	Genotype	Donor
T7	wild type	F.W. Studier
NMλ590	*b538 imm434* $shn6^0$	

2. Phage Techniques (λ and T7.)

a) Plating cells: A fresh overnight bacterial culture was diluted 20-fold in L-broth and grown with aeration at 37°C. When a concentration of 2-5 x 10^8 cells/ml was reached, the cells were pelleted and resuspended in an equal volume of 10 mM $MgSO_4$. The suspension was stored for up to 7 days at 4°C.

b) Titrations: Phage suspensions were serially diluted in phage buffer. 0.1 ml of each dilution was mixed with either 0.2 ml of a suitable indicator strain (plating cell preparation) for λ, or 0.2 ml of a fresh overnight culture for T7. Each infection was mixed with 3 ml BBL Top-agar (at 46°C) containing 10 mM $MgSO_4$, not necessary for T7, and poured onto a BBL plate. Plaques were scored after 12-18 hours incubation at 37°C for λ, or 4-8 hours for T7. Vortex-mixing was never used for infections or dilutions.

c) Plate lysates: Single plaques were picked into 1 ml phage buffer plus 1 drop of chloroform. For λ, 10^6 pfu were mixed with 0.2 ml plating cells and 3 ml BBL Top-agar (at 46°C) containing 10 mM $MgSO_4$, and poured onto a fresh wet L plate. T7 was treated similarly but using 10^3-10^4 pfu per 0.2 mls of a fresh overnight culture. Plates were incubated at 37°C until confluent lysis was achieved; 5 ml L-broth were then added and the plates refrigerated overnight. Broth and top-layer were removed using a bent pasteur pipette, and shaken gently with a few drops of chloroform. The suspension was clarified in a bench top centrifuge at 5000 rpm, 20°C for 10 min, and finally titrated. Titres were in the range 5 x 10^8-10^{11} pfu/ml.

d) Liquid lysates: (Procedures are described for a 100 ml culture, but may easily be scaled up.) A fresh overnight culture was diluted 50-fold in L-broth supplemented with 10 mM $MgSO_4$, and grown at 37°C, with aeration, to OD_{650} 0.45 (2 x 10^8 cells/ml). Phage

were then added with a moi of 1, and the OD_{650} followed until it reached a minimum. Chloroform (0.2 ml) and NaCl (3 g) were then added, shaking continued at 37°C for 10 min and the lysate stood at 4°C for 60 min. The lysate was then cleared at 5000 g, 4°C, 15 min, titrated, and stored at 4°C with 0.2 ml chloroform/ml. Concentration of phage, for DNA preparation, was achieved as described in 2e and 2f.

e) Concentration of phage by Ultracentrifugation: Phage were harvested from clarified lysates by ultracentrifugation for 3 hours at 50,000 g, 4°C. The resulting pellets were resuspended in 0.05 vols phage buffer by gentle rotary shaking overnight at 4°C. Residual debris was removed by a further clarification at 10,000 g, 4°C, 15 min and the supernatant treated at 20°C for 1-3 hours with RNAase A and DNAase I (10 µg/ml of each). Nuclease treatment is not necessary for T7 preparations.

f) CsCl gradient centrifugation: Phage concentrated in the above way were further purified for DNA preparation by one of two procedures. If DNA was required for restriction analysis only, the phage sample ($\geqslant 10^{11}$ pfu) was overlaid onto a CsCl step-gradient. The gradient was prepared in 14 ml polycarbonate centrifuge tubes using 1.5 ml steps of preclarified CsCl solutions of densities 1.3, 1.5 and 1.7 g/ml. The least dense solution was added first and the denser solutions underlaid using a syringe and 21 gauge hypodermic needle. The gradients were run at 33,000 rpm for two hours at 20°C in a 6 x 14 ml MSE Titanium swing-out rotor. Gradients typically showed three bluish bands, the middle of which contained intact phage. This was collected by side-puncture of the tube below the band with a 25 gauge needle. Phage required for storage were also prepared in this way and kept at 4°C. If DNA was required for cloning purposes, samples were further purified

subsequent to the above procedure by CsCl equilibrium centrifugation. The collected band was mixed with 41.5% (w/w) preclarified CsCl solution. The gradients were run in 13 ml polyallomer "quick-seal" tubes (Beckman) using a Spinco 50Ti rotor (48 hours, 38,000 rpm, 20°C). The single phage band was recovered as above.

g) Phenol extraction of phage DNA: Phage collected from CsCl step-gradients or equilibrium gradients were dialysed for 1 hour against TE buffer (10 mM Tris-HCl (pH 7.8), 1 mM EDTA) at 4°C to remove CsCl. The dialysate was extracted four times with an equal volume of freshly-distilled phenol (pre-equilibrated against 0.5 M Tris-HCl (pH 8.0)) at 20°C. Phases were separated by lowspeed centrifugation and the upper DNA-phase removed, keeping well clear of the interface. Residual phenol was removed from the extracted material by extensive dialysis (60 hours) against 8 changes of TE buffer at 4°C. The first buffer batch contained 0.3 M NaCl.. DNA concentrations and purity were estimated by measuring the absorbance at 260 nM, 280 nM and 320 nM. Phage suspensions greater than 10^{13} pfu/ml were diluted before phenol treatment, to avoid inefficient protein extraction.

h) Preparation of λ-phage DNA from plate lysates (Cameron et al, 1977): 10^6 pfu derived from a single plaque were mixed with 0.2 ml plating cells and 3 ml BBL-agarose (10 mM $MgSO_4$) at 46°C, and poured onto a wet L-agarose plate (use of unpurified agar results in poor restriction). After incubation at 37°C until lysis was confluent (5-8 hours), 5 ml of 10 mM Tris-HCl (pH 7.5), 10 mM EDTA were added and the plate left overnight at 4°C. To 4 ml of the supernatant were added 0.4 ml 0.5 M EDTA, 0.2 ml 2 M Tris base, 0.2 ml 10% (w/v) SDS, and 10 µl Diethylpyrocarbonate (sigma). The mixture was heated at 65°C for 30 min in open tubes in the fume cupboard, and

then chilled on ice. 1 ml 5 M potassium acetate (unbuffered) was added, and the solution left on ice for 1 hour before centrifugation (27,000 g, 10 min, 4°C). The clear supernatant was decanted, mixed with 11 ml ethanol in a siliconised Corex tube, and left overnight at -20°C (or for 30 min at -70°C). Nucleic acids were pelleted (12,000 g x, 30 min, 4°C) washed with cold 80% ethanol, dried in a vacuum desicator, and finally dissolved in 200 µl sterile TE buffer.

3. Plasmid Techniques

a) Small scale plasmid DNA preparations for screening purposes (Birnboim and Doly, 1979): 1.5 ml of an overnight culture in L-broth plus antibiotic were centrifuged (2 min, 4°C) in a 1.5 ml "microfuge" snapcap tube using a Quickfit Micro-centrifuge. The cells were washed in 1 ml TE buffer, pelleted, and resuspended in 0.1 ml "Lysis" solution (2 mg/ml Lysozyme, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 50 mM Glucose). After 10 min on ice, 0.2 ml "Alkaline SDS" solution (0.2 M NaOH, 1% (w/v) SDS) was added and the mixture left on ice for a further 10 min, with occasional inversion. A further 0.15 ml of "High Salt" solution (3 M CH_3COONa (pH 5.0)) was added, and the tube was thoroughly mixed by inversion and left on ice for a further 30 min. The resulting heavy white precipitate was pelleted by centrifugation for 6 min at 4°C and the supernatant transferred to a new sterile microfuge tube together with 1 ml of ethanol. The tube was mixed by inversion and placed at -70°C for 30 min to precipitate nucleic acids. These were then pelleted (3 min, 4°C, Quickfit-Microcentrifuge), resuspended in 0.05 ml 0.3 M CH_3COONa (pH 5.0) and reprecipitated by the addition of 0.125 ml of ethanol and incubation at -70°C for a further 30 min. Nucleic acids were once more pelleted, dried in a vacuum desicator, and

redissolved in 0.05 ml of TE buffer. This method proved extremely useful when screening recombinants, as 24 samples can easily by analysed in one day using 10-30 µl of final DNA solution per restriction endonuclease digest. Should it be necessary, RNA can be removed from this preparation by the addition of 1 µl RNAase A (1 mg/ml, pre-treated for 15 min at 95°C) and incubation at 37°C for 30 min.

b) Large-scale preparation of plasmid DNA: A single colony was inoculated into 100 ml L-broth supplemented with selective antibiotic, and incubated at 37°C with aeration for 16-18 hours. The cells were pelleted (5000 g, 15 min, 4°C) and washed with TE buffer. Unless otherwise stated, all the following treatments and additions were at 2-4°C. The cells were resuspended in 1.5 ml 25% (w/v) sucrose solution and treated with 0.5 ml lysozyme solution (10 mg/ml). The mixture was left for 5 min prior to the addition of 0.5 ml 0.5 M EDTA (pH 8.5) and 0.2 ml RNAase A (5 mg/ml, pretreated at 95°C for 15 min). A further 5 min interval preceded the addition of 2.5 ml "Triton" solution (per 100 ml: 1 ml 10% (w/v) Triton-X-100, 12.5 ml 0.5 M EDTA (pH 8.5), 5 ml Tris-HCl (pH 8.0)). The mixture was gently inverted for 10 min and then cleared by centrifugation (27,000 g, 20 min). The resulting supernatant was made up to 9 ml with TE buffer and 9 g CsCl plus 0.9 ml ethidium bromide (5 mg/ml) were added. The solution was then centrifuged (50Ti rotor, 90,000 g 15°C, for 60 hours) and the tube viewed under long-wave UV. Typically, two fluorescent bands could be seen, the lower of which contained supercoiled plasmid DNA. This was collected by side-puncture of the tube below the band using a syringe and 19 gauge needle. The preparation was dialysed against TE buffer to remove CsCl, and extracted twice with an equal volume of freshly-distilled

phenol (pre-equilibrated against 0.5 M Tris-HCl (pH 8.0)) at 20°C. Residual phenol was removed from the extracted material as for phage DNA preparations, and the DNA was collected by ethanol precipitation. DNA concentrations and purity were estimated by measuring the absorbance at 260 nM, 280 nM and 320 nM. This method gave yields of 1 to 7.5 mg/litre for ColE1 derivatives, and a minimum of 0.5 mg/litre for lower-copy-number plasmids.

4. DNA Techniques

All DNA solutions were stored at 4°C.

a) Ethanol precipitation: DNA was precipitated by the addition of 0.1 vol 3 M potassium acetate (pH 6) and 2-3 vol ethanol, followed by incubation at -20°C (12 hours) or -70°C (30 min), in siliconised Corex tubes or plastic microfuge tubes. Nucleic acids were pelleted (12,000 g, 30 min, 4°C, or 3 min in a Quickfit micro-centrifuge, 4°C), washed with 80% ethanol, dried in a vacuum desicator and redissolved in TE buffer.

b) Restriction endonuclease digestions: Appropriate quantities of DNA, 10x reaction buffer, dH_2O and restriction endonuclease were mixed thoroughly in plastic microfuge tubes and incubated at appropriate temperatures, typically for 1 hour. Reactions were stopped by the addition of excess EDTA and/or incubation at 70°C for 10 min. Table 2.3 lists the enzymes used and the reaction conditions for each.

c) DNA fragment purification: Fragments were purified from agarose gels by passive elution. The required region of gel was excised and homogenised by passage through a 19 gauge needle into a siliconised Corex tube. After the addition of 5 ml PE buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM EDTA), the suspension was shaken at 37°C for 12-24 hours. Agarose fragments were removed by

TABLE 2.3 RESTRICTION ENDONUCLEASES

Endonuclease	Bacterium of Origin	Specificity	Source	Assay Conditions
AluI	Arthrobacter luteus	AG▼CT	A Newman*	10 mM Tris-HCl (pH 7.6); 7 mM $MgCl_2$; 6 mM 2-mercaptoethanol
AvaI	Anabaena variabilis	C▼PyCGPuG	J C Boothroyd*	6 mM Tris-HCl (pH 7.5); 6 mM $MgCl_2$; 6 mM NaCl, 6 mM 2-mercaptoethanol
BamHI	Bacillus amyloliquefaciens H	G▼GATCC	BRL	20 mM Tris-HCl (pH 7.0); 7 mM $MgCl_2$; 100 mM NaCl; 2 mM 2-mercaptoethanol
EcoRI	Escherichia coli RY13	G▼AATTC	BOEHRINGER	10 mM Tris-HCl (pH 7.5); 10 mM $MgCl_2$; 50 mM NaCl; 6 mM 2-mercaptoethanol
HaeIII	Haemophilus aegyptius	GG▼CC	NEB	6 mM Tris-HCl (pH 7.4); 6.6 mM $MgCl_2$; 6 mM NaCl; 6 mM 2-mercaptoethanol
HindIII	Haemophilus influenzae Rd	GTPy▼PuAC	Mrs K Mileham*	6.6 mM Tris-HCl (pH 7.4); 6.6 mM $MgCl_2$; 1 mM dithiothreitol
HindIII	Haemophilus influenzae Rd	A▼AGCTT	NEB	7 mM Tris-HCl (pH 7.4); 7 mM $MgCl_2$; 60 mM NaCl; 6 mM 2-mercaptoethanol
HpaI	Haemophilus parainfluenzae	GTT▼AAC	I Garner	10 mM Tris-HCl (pH 7.4); 10 mM $MgCl_2$; 6 mM 2-mercaptoethanol
KpnI	Klebsiella pneumoniae OK8	GGTAC▼C	BRL	6 mM Tris-HCl (pH 7.5); 6 mM $MgCl_2$; 6 mM 2-mercaptoethanol

centrifugation and passage of the supernatant through siliconised glass wool. The suspension was passed twice through a 0.1 ml DEAE-cellulose (Whatman, DE52) column in a 1 ml plastic syringe at 5 ml hr^{-1}. DNA bound to the column was eluted with 0.5 ml PEN buffer (50 mM Tris-HCl (pH 8.0), 1 M NaCl, 10 mM EDTA) and concentrated by ethanol precipitation. The method gave good yields (50-90%) and its products were suitable for restriction, ligation and nick-translation.

d) Recombination of DNA fragments *in vitro*: Restriction endonuclease digests containing DNA fragments possessing cohesive termini were diluted to 5-30 µg DNA/ml with 10 mM Tris-HCl (pH 7.5), 100 mM NaCl. 0.1 vol of 10x T4 ligase cocktail (660 mM Tris-Cl (pH 7.2), 10 mM EDTA, 100 mM $MgCl_2$, 100 mM Dithiothreitol, 1 mM ATP) were added together with T4 DNA ligase. Ligation was at 10°C for 3-12 hours, after which samples were withdrawn for transfection/transformation. Digests containing DNA fragments possessing flush termini were treated similarly but with the following differences: i) concentrations of 200-500 µg/ml DNA were used, ii) 10x ligation cocktail was 10 mM ATP, iii) 10-fold more T4 DNA ligase was used per reaction.

e) Transfection/transformation: A fresh overnight culture was diluted 50-fold in L-broth and incubated at 37°C, with aeration, until an OD_{650} of 0.5-0.6 was reached. The material was transferred to an ice-bath for the remaining procedures. The cells were pelleted (5,000 g, 10 min, 4°C), washed in 0.5 vol TE buffer, and resuspended in 0.5 vol 0.1 M $MgCl_2$. The cells were then pelleted, resuspended in 0.5 vol 0.1 M $CaCl_2$ and left at 2°C for 20 min. After a final pelleting, the cells were resuspended in 0.05 vol 0.1 M $CaCl_2$ and left for 1 hour before use. DNA was diluted

in 0.1 ml SSC:$CaCl_2$ (3 vol 1x SSC: 4 vol 0.1 M $CaCl_2$) and 0.2 ml competent cells were added. After 30 min the cells were heat-shocked at 42°C for 2 min, and returned to ice for a further 30 min. For transfections, the cells were then mixed with 3 ml BBL top-agar (46°C), supplemented with 10 mM $MgSO_4$, and plated out. For transformations, 1 ml L-broth was added and the mixture incubated at 37°C for 1 hour before plating on selective medium.

f) Use of synthetic molecular recombination linkers: The method was essentially as described by Maniatis et al (1978). 2 µg HindIII molecular recombination linker (dCCAAGCTTGG from Collaborative Research Inc.) were phosphorylated using T4 polynucleotide kinase (10 units/µl, 24 units total) in a total volume of 10 µl PNK buffer (20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM ATP, 10 mM ß-mercaptoethanol) for 1 hour at 37°C. The reaction mix was then added directly to 8 µg AluI digested T7 DNA (in 10 µl PNK buffer) and 1 µl T4 DNA ligase. The mixture was incubated at 10°C for 4 hours, followed by 70°C for 10 min, and returned to ice. The efficiencies of the phosphorylation/blunt-end ligation steps were monitored by running one-tenth of the mixture on a 3-10% polyacrylamide gradient gel and examining the ethidium-stained gel under UV (see 1/iv). A successful phosphorylation/ligation was evidenced by the presence of a ladder of linker oligomers (2 to 15-mers). The remaining mixture was digested with 40 units HindIII for 1 hour at 37°C. Linker fragments were then removed by passage through a Sephadex G-75 (Pharmacia) column (8 cm x 0.5 cm). The column was eluted with TE buffer and 10 x 200 µl fractions collected. 50 µl samples were assayed for DNA fragments $>$ 50 bp long by gel electrophoresis (3-10% polyacrylamide). Fractions containing such DNA were concentrated by ethanol precipitation and used in ligations

with HindIII digested λ590 DNA. Following transfection, AluI fragments with synthetic HindIII cohesive termini were recovered in λ-recombinants exhibiting a clear plaque morphology (due to the insertional inactivation of the imm434 repressor-gene).

g) ^{32}P-labelling of DNA by nick translation (Rigby et al, 1977): 10 μCi α^{32}-P-dCTP (400 Ci/mMole) was dried in a vacuum desicator, washed twice with 20 μl dH_2O, and resuspended in 20 μl of 50 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 20 μM dATP, 20 μM dGTP, 20 μM dTTP containing 1 μg of the appropriate DNA. The reaction was started by addition of 1 μl DNAase I (10^{-4} mg/ml) and 1 μl E.coli DpolI (1 unit/μl) and incubated at 15°C. Incorporation of label was followed by monitoring TCA-precipitable counts bound to Whatman GF/C filters (2.4 cm). The reaction was stopped by the addition of 0.2 ml TNE buffer (10 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1 mM EDTA) when incorporation ceased. A drop of Orange-G marker dye was added, and unincorporated nucleotides removed by passage over a Sephadex G-50 column (20 cm x 1 cm). 1 ml fractions were eluted with TNE buffer until the Orange-G (which co-migrates with dNTPs) had reached the bottom of the column. Peak fractions were identified by Cerenkov counting, and stored at -20°C.

h) Terminal labelling of nucleic acid molecules (Murray, 1973): DNA for labelling (1-5 μg) was first treated with 1 μl Calf Intestinal Phosphatase (15 units/μl) in TE buffer or restriction buffer, for 30 min at 37°C. The reaction was stopped by the addition of one-tenth vol 0.5 M EDTA and incubation at 70°C for 15 min. The DNA was then ethanol-precipitated, washed in 80% ethanol, dried in a vacuum desicator, and resuspended in 10 μl 1.2 x PNK buffer (1 x PNK buffer was 20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM β-mercaptoethanol). Sufficient γ-^{32}P-ATP (3000 Ci/mMole)

for the labelling reaction was supplemented with cold ATP and dried in a vacuum desicator. (A final ATP concentration of 1 μM is required to label single-stranded 5'-ends, whereas 10-20 μM is required to label blunt or recessed ends.) The DNA solution was used to redissovle the γ-^{32}P-ATP, and 2 μl T4 polynucleotide kinase (10 units/μl) added to start the reaction. (When labelling blunt or recessed ends, the DNA should be denatured by incubation at 90°C for 2 min and quenching in ice, immediately before adding the ATP.) The mixture was incubated at 37°C for 1-2 hours, and the reaction terminated by incubation at 70°C for 15 min. The mixture was phenol-extracted once, ethanol-precipitated twice, washed in 80% ethanol twice, and resuspended in TE buffer.

i) Plaque detection by DNA:DNA hybridisation (Benton and Davis, 1977): A circular nitrocellulose filter (Millipore, 8 cm diameter) was placed carefully onto each plate carrying plaques, and left for 2 min. The filter was then carefully removed and placed on the following, plaque side uppermost: Whatman 3 MM paper saturated with:

i) 0.5 M NaOH (5 min)

ii) 0.1 M NaOH, 1.5 M NaCl (5 min)

iii) 0.5 M Tris-HCl (pH 7.5), 1.5 M NaCl (5 min)

iv) 0.5 M Tris-HCl (pH 7.5), 1.5 M NaCl (5 min)

The filters were blotted dry, baked for 2 hours at 80°C in vacuo, and then prehybridised in hybridisation solution (0.02% (w/v) Ficoll 400, 0.02% (w/v) polyvinylpyrrolidone, 0.02% (w/v) gelatin, 0.1% (w/v) SDS, 4x SSC, 50% formamide, 50 μg/ml denatured, sonicated Salomon sperm DNA) at 37°C for 2-4 hours. The filters were then transferred to a sealed polythene bag containing 10-20 ml Hybridisation fluid and $\geq 5 \times 10^5$ cpm ^{32}P-labelled probe DNA (previously denatured by incubation at 90°C for 10 min). Filters were

hybridised in this solution for 12-18 hours at 37°C, then washed in 4 x 500 ml 2 x SSC, 0.5% (w/v) SDS, and dried at 37°C. Autoradiography was at -70°C using flash-sensitised X-ray film (DuPont, Cronex 4) and an intensifying screen (Ilford Fast Tungstate), usually for 12-48 hours.

j) Colony Detection by DNA:DNA hybridisation (Grunstein and Hogness, 1975): Colonies of plasmid-containing bacteria were transferred by sterile toothpick, in an ordered array, to an L-agar plate containing selective antibiotic. A circular nitrocellulose filter (8 cm diameter, Millipore) was carefully placed on the surface of the plate, which was then incubated overnight at 37°C. The filter was removed and placed on the following, colony side uppermost: Whatman 3MM paper saturated with:

i) 0.5 M NaOH (7 min)

ii) 1 M Tris-HCl (pH 7.4) (2 min)

iii) 1 M Tris-HCl (pH 7.4) (2 min)

iv) 0.5 M Tris-HCl (pH 7.4), 1.5 M NaCl (4 min)

The filter was blotted dry, baked for 2 hours at 80°C *in vacuo*, and subsequently treated as above (see 4/i).

k) DNA fragment detection in agarose gels by DNA:DNA hybridisation (Southern, 1975): Agarose gels were stained with ethidium bromide and photographed to visualise DNA fragments. Regions of the gel of interest were immersed in the following:

i) 0.5 M NaOH, 1.5 M NaCl (40 min)

ii) 0.5 M Tris-HCl (pH 7.4), 3 M NaCl (40 min)

Treated sections of gel were placed on 2 sheets of Whatman 3 MM paper (presoaked in 20 x SSC), supported over a reservoir of 20 xSSC. The edges of the paper dipped into the reservoir and functioned as wicks. An exact gel sized piece of nitrocellulose (Millipore), soaked

in 2 x SSC, was placed over the gel, followed by a similarly sized piece of Whatman 3 MM paper soaked in 2 x SSC. Throughout, air-bubbles were excluded at interfaces. A 10 cm stack of similarly sized dry Whatman 3 MM paper was placed on top, followed by a suitably sized glass-plate. The construction was covered with Vitafilm (Goodyear) and a weight (usually a 5 l plastic container of Mineral oil) placed on top. Overnight, capillary action drew SSC from the reservoir, through the gel, through the nitrocellulose filter to the paper above. Denatured DNA carried with the SSC was retained on the nitrocellulose filter, which was then washed in 2 x SSC, blotted dry and baked at 80°C for 2 hours in vacuo. The filter was subsequently treated as above (see 4/i) to highlight DNA fragments of interest.

1) Gel electrophoresis:

(i) Horizontal agarose mini-gels:- Gels were prepared by boiling 0.2 g agarose in 25 ml TAE buffer (40 mM Tris-Acetate (pH 8.3), 33 mM sodium acetate, 1 mM EDTA) with vigorous stirring, on an electric hot-plate/magnetic stirrer for 30 sec. The agarose solution was pipetted onto a horizontal glass plate (7 cm x 7 cm x 0.2 cm) over which a seven-tooth slot-former was supported (1 mm clearance) with bulldog clips. The agarose solution, retained on the plate by surface-tension, set within 5 min. The slot-former was removed, the plate supported over two reservoirs containing 400 ml TAE buffer, and pieces of Jay-cloth hung from each end to make electrical contact with the reservoirs. DNA samples for analysis (10 µl) were mixed with 5 µl TAE loading buffer (0.5 x TAE, 0.125 M EDTA, 0.1% (w/v) bromophenol blue, 50% (w/v) glycerol) and loaded by micropipette. After applying 50V for 5 min, the surface was wetted with TAE buffer and covered with Vitafilm, and the voltage increased to 125V for 30 min.

DNA (0.1-0.5 g) was visualised by staining with ethidium bromide (Sigma, 2 μg/ml) for 10 min, destaining in dH_2O for 2 min, and transillumination with long-wave UV (Ultraviolet Products Inc., Chromatovue C-62, 365 n. peak transmission). Photography was with Ilford FP4 film using 20 sec exposure through a red filter.

(ii) Horizontal agarose gels (McDonnell et al, 1977):- Gels (1%-1.5% (w/v)) were prepared by boiling agarose in 200 ml TAE buffer with vigorous stirring, on an electric hot plate/magnetic stirrer for 1 min. The agarose solution was then cooled to 50°C and poured into a perspex mould (30 cm x 14 cm x 0.5 cm) sealed at each end with perspex spacers (retained by bulldog clips) and having a 13-tooth slot-former in place 5 cm from one end. After the agarose had set, the slot-former and end-spacers were removed, and the gel was prepared for electrophoresis as above. DNA samples for analysis (30 μl) were mixed with 5 μl TAE loading buffer and loaded by micropipette. Gels were run at 50V for 15 min, the surface wetted with TAE buffer and covered with Vitafilm, and the voltage increased to 100V for 16-20 hours. DNA (0.5-2 μg) was visualised as in 1/i, except ethidium bromide staining was for 30 min and destaining was for 45 min.

(iii) Vertical 5% polyacrylamide gels:- Stock solutions for these gels were: 30% acrylamide (29% (w/v) acrylamide, 1% (w/v) bisacrylamide)

10 x TBE buffer (0.9 M Tris base, 0.89 M boric acid, 25 mM EDTA, pH 8.3)

10% (w/v) ammonium persulphate in dH_2O (freshly prepared)

A 40 ml solution containing 6.625 ml 30% acrylamide, 4 ml 10 x TBE, 0.5 ml 10% ammonium persulphate, and 28.88 ml dH_2O was degassed (2 min swirling on a waterpump). 25 μl TEMED (N,N,N',N'-tetramethylethylenediamine, Serva) were added and the solution poured

into a glass sandwich (20 cm x 16 cm x 0.15 cm) with perspex spacers sealed with water-agar. The gel was overlaid with 2-butanol and allowed to polymerise ($\leqslant$1 hour). A stacking gel mixture containing 1.67 ml 30% acrylamide, 2 ml 10 x TBE, 0.33 ml ammonium persulphate and 16 ml dH_2O was prepared and degassed while the gel was setting. When polymerisation was complete, the 2-butanol was decanted and the surface of the gel rinsed in 1 x TBE buffer. Following the addition of 10 µl TEMED, the stacking gel was overlaid by pipette, a 13-tooth slot-former inserted, and the surface areas covered with 2-butanol. After polymerisation, the slot-former and bottom spacer were removed and the gel attached to a perspex gel apparatus (Raven), with 400 ml 1 x TBE buffer in each reservoir. DNA samples for analysis were ethanol-precipitated, dried in a vacuum desicator, and resuspended in 20 µl TE buffer. After the addition of 10 µl TBE loading buffer (0.5 x TBE, 0.125 M EDTA, 0.1% bromophenol blue, 50% (v/v) glycerol), samples were dried to 10-15 µl prior to loading with a Hamilton microsyringe. Gels were run at 300V for 1 hour and DNA (2-4 µg) was visualised as in 1/i.

(iv) Vertical 3-10% gradient polyacrylamide gels (Maniatis et al, 1975):- Stock solutions for these gels were as in 1/iii. Two 30 ml solutions containing 3 ml and 10 ml 30% acrylamide respectively, 3 ml 10 x TBE, 0.25 ml 10% ammonium persulphate and dH_2O to 30 ml, were degassed as before. After addition of 10 µl TEMED to each, a 2-chamber linear gradient-maker was used to pour a 3-10% gradient into a glass sandwich (40 cm x 1 cm x 0.15 cm) with perspex spacers sealed with water-agar. The gradient was overlaid with 2-butanol and allowed to polymerise. A stacking gel mixture containing 1.67% ,; acrylamide, 2 ml 10 x TBE, 0.33 ml ammonium persulphate and 16 ml dH_2O was prepared and degassed while the lower gel was setting.

The 2-butanol was then decanted, the gel surface rinsed with 1 x TBE, 10 µl TEMED added to the stacking gel and the latter overlaid by pipette. A 13-tooth slot-former was then inserted and the surface areas covered with 2-butanol. After polymerisation, the gel was set up and samples applied as in 1(iii), above. Electrophoresis was at 250V for 30 min, followed by 120V for 15 hours. DNA (2-4 µg) was visualised by ethidium bromide staining (as in 1/i) and gels were photographed against a black background by overhead shortwave UV illumination (Mineralight C5, 254 nM peak transmission) using 5-8 min exposure through a red filter, and Ilford FP4 film.

m) Gel autoradiography: Gels were dried onto Whatman 3 MM paper (BIORAD Laboratories Gel Slab Drier) and autoradiographed at -70°C using flash-sensitised X-ray film (DuPont, Cronex 4) and an intensifying screen (Ilford Fast Tungstate), usually for 12-48 hours. Sequencing gels were wrapped in Saranwrap, autoradiographed wet at room temperature using X-ray film (DuPont, Cronex 4) for 12-48 hours.

5. Enzyme assays: Assays of two enzymes were used.

a) Assay of β-galactosidase (Miller, 1972):- cultures were grown at 37°C on an orbital shaker at 200 rpm in Spizizen minimal medium containing ampicillin (50 µg/ml), acid casein hydrolysate (0.1% (w/v)), and glycerol (0.2% (w/v)) to avoid catabolite repression. β-galactosidase levels were determined from samples taken at OD_{650} between 0.2-0.6. Each assay contained the following: 0.8 ml PM2 reducing buffer (120 mM sodium phosphate (pH 7.0), 1 mM $MgSO_4$, 0.2 mM $MnSO_4$, 10 mM β-mercaptoethanol); 0.2 ml 0.1% (w/v) CTAB; 0.01 ml 1% (w/v) sodium deoxycholate; 1 ml of a suitable dilution of cells; and 0.6 ml 13.5 mM o-nitrophenyl β-D-galactopyranoside (Sigma). Assays were incubated at 28°C and were stopped after sufficient colour had developed (usually 20-100 min) by the addition of 1.3 ml

1 M Na_2CO_3. The A_{550} and A_{420} were measured using a Zeiss spectrophotometer and β-galactosidase units calculated as follows:

$$\text{Miller Units (per ml of culture at } OD_{600} \text{ of 1.00)} = \frac{100 \times OD_{420} - (1.75 \times OD_{550})}{t \times V \times OD_{600}}$$

where V is the volume of culture in sample (ml)

t is the duration of 28°C incubation in minutes

OD_{600} is the culture absorbance at time of sampling.

b) Assay of galactokinase (McKenney et al, 1981):- Cultures were grown at 37°C on an orbital shaker at 200 rpm in Spizizen minimal medium containing ampicillin (50 µg/ml), acid casein hydrolysate (0.1% (w/v)) and fructose (0.2% (w/v)) to avoid catabolite repression. Galactokinase levels were determined from samples taken at OD_{650} 0.2. 1 ml samples were treated with 40 µl of lysis buffer (100 mM EDTA, 100 mM Dithiothreitol, 50 mM Tris-HCl (pH 8.0)) and two drops of toluene. Lysates could be stored on ice for at least 60 min prior to the evaporation of toluene using an air-line (20°C) and assay. Each assay contained the following: 20 µl 5 mM Dithiothreitol/16 mM NaF, 50 µl 8 mM $MgCl_2$/200 mM Tris-HCl (pH 7.9)/ 3.2 mM ATP, 10 µl 10 mM D-galactose (containing ^{14}C-galactose to give a final specific activity of 4.5×10^6 dpm/µmole; Amersham), and 20 µl lysed culture. Assays were incubated at 32°C for 10-30 min and 50 µl aliquots removed to 2.5 cm discs of DE81 paper (Whatman). Filters were washed in dH_2O (Hayward, 1980) together with two blank control filters. Total counts were determined by removing two random 25 µl aliquots to DE81 discs, these remaining unwashed. Filters were dried at 80°C, counted by scintillation in 0.4% butyl-PBD (GBA)/toluene (Packard TRI-CARB scintillation spectrophotometer), and galactokinase units calculated as follows:

$$\text{Galactokinase units (nanomoles galactose phosphorylated per min per } OD_{650}) = \frac{(cpm - B) \times 2500}{U \times t \times OD_{650}}$$

where U is the average count on unwashed filters

B is the average count on blank filters

t is the duration of 32°C incubation in minutes

OD_{650} is the culture absorbance at time of sampling

6. M13 sequencing: This method has been extensively discussed in the literature and will not be described here (Sanger et al, 1977; Sanger and Coulson, 1978; Smith, 1974; Schreier and Cortese, 1979; Anderson et al, 1980; Messing et al, 1981). However, the principle of the method will be discussed briefly later (see Fig. 5.1.).

CHAPTER 3 - ISOLATION OF T7 DNA FRAGMENTS

3.1 Introduction

The study of transcriptional terminators in their native environment has three major disadvantages: simple screening procedures to monitor terminator efficiency are usually not available; directed mutagenesis specifically affecting terminator sequences is difficult to achieve using genomic DNA; isolation of mutant sequences for characterisation is relatively tedious. Moreover, if each regulatory site is studied in context, valid comparisons between their efficiencies and factor dependencies are impossible as there is no standard for comparison. The technique of gene fusion has proved extremely useful in overcoming some of these problems. Genetic fusions can place a monitor gene, whose function is readily assayed and is subject to powerful genetic selection techniques (e.g. *lac*, *gal* or *trp* genes), under the control of different regulatory elements. This allows comparative study of transcriptional signals in a single genetic environment. For example: *in vitro* recombination can be used to insert a transcriptional terminator between a given promoter and a suitable monitor gene. The vector should possess an independent, selectable marker and facilitate the assay of its monitor gene function *in vivo*. It is desirable that it should also allow: the transfer of mutant or wild-type fragments into other suitable vectors; the characterisation of terminators by *in vitro* techniques (e.g. DNA sequencing and transcription studies); the easy recovery of terminator fragments to permit specific mutagenesis *in vitro*; and the forced reconstruction of the original fusion using mutant DNA. It is, therefore, advantageous if the cloned fragments are determined by cohesive termini specified by one (or preferably two) of the more commonly used restriction endonucleases.

The ara-trp-lac fusion (Casadaban, 1975) present on the ColE1-derived plasmid vector, pMC81, allows the construction of recombinants of this type (Casadaban and Cohen, 1980). A derivative of this vector, pHR1, has lacZ expression dependent on transcription initiating at the constitutive promoters P_{L10} and P_{L11}, from the rplKAJL gene cluster of E.coli (Newman, 1980). The derivative contains a single site for HindIII, at which DNA fragments can be inserted between the ribosomal gene promoters and the lacZ structural gene (Fig. 3.1). The following work attempted to exploit pHR1 to study transcriptional terminators derived from the T7 genome; in particular, tec1.

3.2 Construction of T7 derived recombinants

The construction of such recombinants poses special problems in the case of T7 DNA. The phage genome does not contain the DNA sequences recognised by the more commonly used restriction endonucleases (e.g. EcoRI, HindIII, PstI, BamHI, SalI) (Rosenberg et al, 1979; Dunn and Studier, 1982). Presumably the lack of cleavage sites for these enzymes is due to their elimination by natural selection during phage infection of various hosts. Furthermore, the remaining suitable enzymes generate relatively large DNA fragments which present two disadvantages. Firstly, they are likely to contain a lethal element from the T7 genome such as a strong promoter or a lethal gene. Secondly, the presence of multiple transcriptional signals and/or translational properties which can distort transcription is a likely source of difficulties.

The strategy adopted was, therefore, to fragment T7 DNA with restriction endonucleases that produce small fragments with flush-ends, one of which should contain tec1 and little else. The

Fig. 3.1

The upper part shows the genetic organisation and physical map of pMC81, linearised at one of its 3 BamHI sites (Casadaban and Cohen, 1980). Transcription initiates in the araOI region, preceeding leftwards into araC, and rightwards through araB', 'Mu', trp'CBA' into 'lacZ. The fusion of trpA' to 'lacZ creates a hybrid, functional protein that is noticeably temperature sensitive (Squires, 1982). The co-ordinates of restriction sites are marked in kilobase-pairs starting from the BamHI site lying between rep ColE1 and araC.

The lower part shows, at an expanded scale the genetic organisation and physical map of the first 11 kb of pHR1 (Newman, 1980). The total size of the vector is 27.4 kb. Transcription initiates in the rpo P_{L11} P_{L10} region and proceeds rightwards into 'lacZ through trp'BA'. The co-ordinates of restriction sites are given in kilobase-pairs except for 4 PstI sites between the KpnI and BglII sites which have been omitted for clarity.

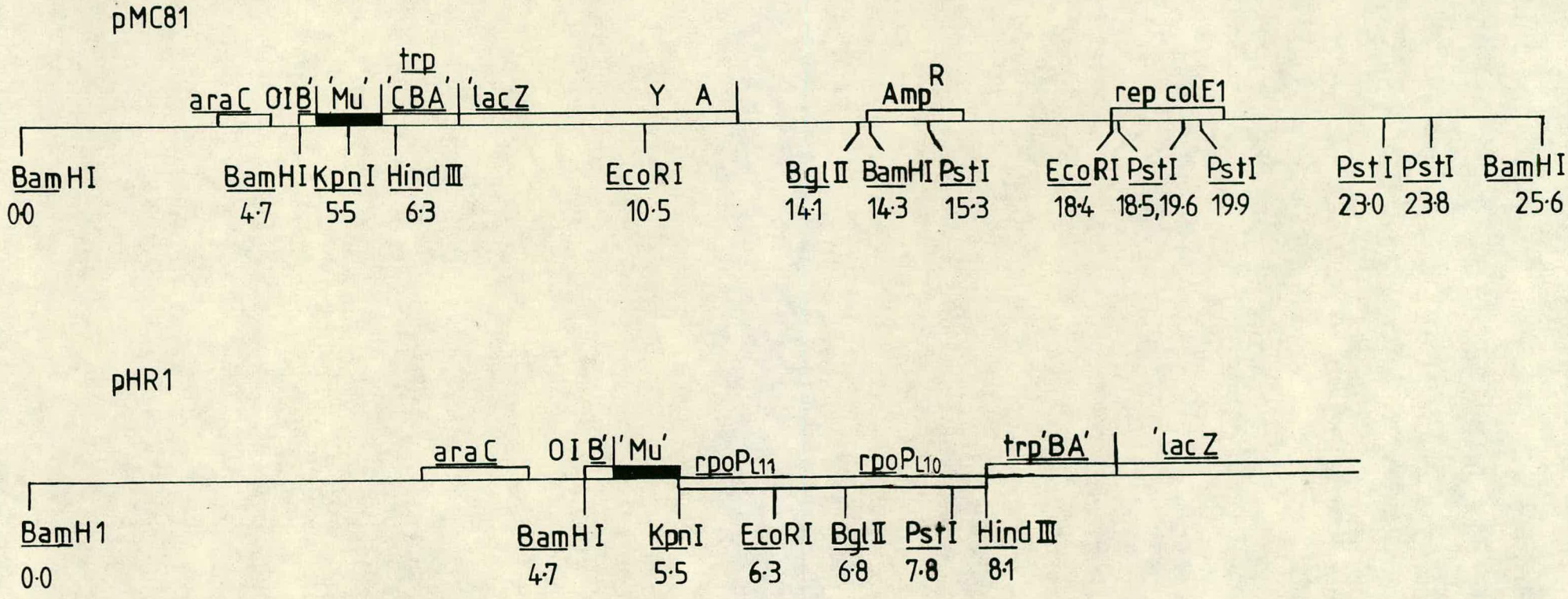

pMC81
trp
araC OIB' 'Mu' 'CBA' 'lacZ Y A
Amp R
rep colE1
BamHI
0·0
BamHI KpnI HindIII
4·7 5·5 6·3
EcoRI
10·5
BglII BamHI PstI
14·1 14·3 15·3
EcoRI PstI PstI
18·4 18·5,19·6 19·9
PstI PstI BamHI
23·0 23·8 25·6
pHR1
araC OIB' 'Mu' rpoPL11 rpoPL10 trp'BA' 'lacZ
BamH1
0·0
BamHI KpnI EcoRI BglII PstI HindIII
4·7 5·5 6·3 6·8 7·8 8·1

Fig 3·1

enzymes chosen were AluI, HpaI and HindII; the former generates a 184 bp tec1 fragment (T7 co-ordinates 18.68% to 19.12%), and the latter two the same 446 bp tec1 fragment (T_7 co-ordinates 18.27% to 19.40%). The resulting fragments were provided with HindIII cohesive termini by the addition of synthetic molecular linkers and digestion with HindIII (Fig. 3.2). DNA treated in this way was ligated to HindIII-digested pHR1, and transformed into MC1000. The resulting colonies were screened by colony-hybridisation with a [^{32}P]-3575 bp KpnI fragment (KpnI D, T7 co-ordinates 14.04% to 22.98%) for fragments originating from the region of the T7 genome in which tec1 lies. Transformants carrying plasmids were selected on lactose-MacConkey agar containing ampicillin. This medium allowed the secondary screening of all colonies for a reduction in lacZ expression, as judged by their colour phenotype. However, this procedure failed to identify a pHR1 derivative carrying either tec1 or any other fragment producing a reduction in lacZ expression.

This prompted the use of phage λNM590, in which HindIII-insertions inactivate immunity functions, as an initial vector (Murray et al, 1977). This has three major advantages over pHR1 for the preliminary isolation of recombinants: (i) a high transformation efficiency enables the recovery of many recombinants; (ii) the clear-plaque phenotype of recombinants simplifies their identification; (iii) plaque-hybridisation is more sensitive than colony-hybridisation as a means of identifying desired recombinants.

T7 DNA was, therefore, treated as above with AluI and synthetic molecular linkers, but ligated to HindIII-digested λNM590 (Fig. 3.2). Note that AluI generates ≥90 fragments from T7 DNA (Fig. 3.3). Following transfection into ED8654, recombinants were initially identified by their clear-plaque morphology, plaque-purified,

Fig. 3.2

Flow diagram summarising the final strategy adopted for cloning the 184 bp tec1 AluI fragment. Further details can be found in sections 2.4.f and 3.2.

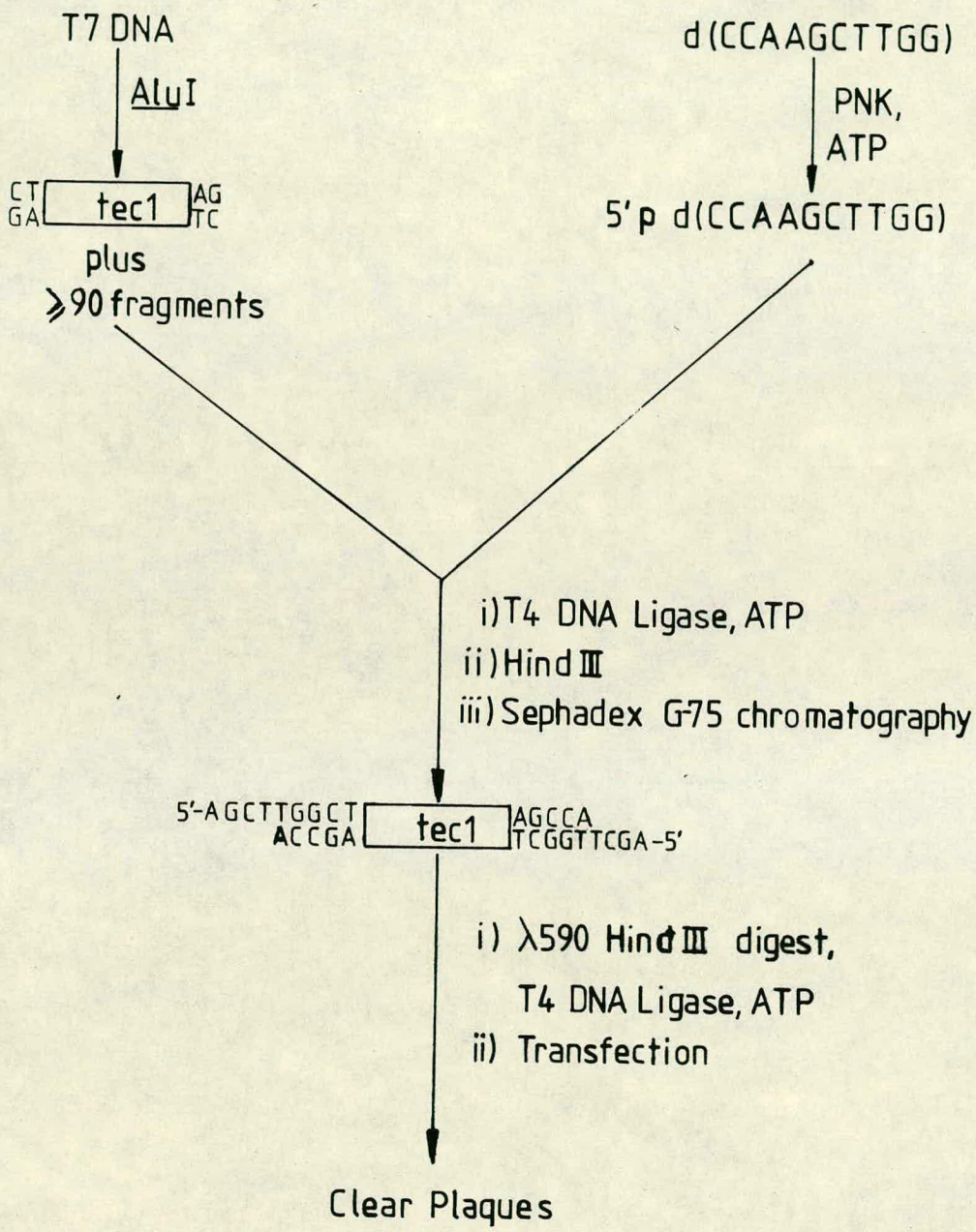

T7 DNA
AluI
CT
GA
tec1
AG
TC
plus
⩾90 fragments
d(CCAAGCTTGG)
PNK,
ATP
5'p d(CCAAGCTTGG)
i) T4 DNA Ligase, ATP
ii) Hind III
iii) Sephadex G-75 chromatography
5'-AGCTTGGCT
ACCGA
tec1
AGCCA
TCGGTTCGA-5'
i) λ590 Hind III digest,
T4 DNA Ligase, ATP
ii) Transfection
Clear Plaques

Fig 3·2

and transferred by cocktail stick in an ordered array onto a lawn of ED8654. Plaques containing phage carrying fragments of T7 DNA were identified by plaque-hybridisation of this array with a [^{32}P]-T7 DNA probe, radiolabelled by nick-translation. In this way, 93 recombinant phage were identified, exhibiting a varying degree of homology with the T7 probe as judged by "spot" intensity after autoradiography. (This correlation can only be considered quantitative if we assume: equal distribution of ^{32}P-label in the probe; similar concentrations of phage within plaques; efficient transfer of all phage DNA to nitrocellulose filters.)

3.3 Screening of the T7 library for tec1

The array was next plaque-hybridised to a [^{32}P]-3575 bp KpnI fragment D, as in previous colony-hybridisations, to identify phage carrying fragments originating from the region of the T7 genome in which tec1 lies. Sequence data predicted that this probe should hybridise to 13 T7 AluI fragments; among these the tec1 fragment is 184 bp, and no others are close in size (Dunn and Studier, 1980). The desired clone should, therefore, have been easily identified if present. The probe highlighted 11 phage, which were purified on CsCl step-gradients and phenol-extracted to yield DNA suitable for restriction analysis. Digestion with HindIII and analysis on 3-10% gradient polyacrylamide gels revealed that the majority of these phage carried multiple HindIII-inserts of varying sizes (Table 3.1). Further characterisation by AluI digestion revealed that several of the larger HindIII fragments contained internal AluI sites (Table 3.1); these were presumably the products of the T7 AluI-fragment ligation. It is unlikely that they arose from incomplete AluI digestion as digestions used in these experiments

Table 3.1

The recombinant phage hybridising to T7 KpnID are recorded. The table indicates the sizes of the HindIII inserts they contain (in bp) and whether or not they contain an internal AluI site (+ or -). Larger fragments of limited interest were not tested (N.T.) as they probably arose from several T7 AluI fragments. The sizes of the products of HindIII/AluI digestion are recorded where observed.

TABLE 3.1 Recombinant Phage Hybridising to T7 KpnI-D

λIG	Size of insert/s	Internal AluI sites	AluI Products
9	800 bp	+	not detected (n.d.)
	650 bp	+	(n.d.)
10	720 bp	+	655 bp and 65 bp
	600 bp	-	
17	305 bp	-	
	220 bp	+	(n.d.)
21	330 bp	-	
36	1275 bp	N.T.	
46	310 bp	-	
	260 bp	-	
	198 bp	-	
74	600 bp	+	(n.d.)
	281 bp	+	232 bp and 49 bp
	194 bp	-	
	174 bp	-	
77	1375 bp	N.T.	
	1300 bp	N.T.	
79	720 bp	-	
	188 bp	-	
81	1150 bp	+	(n.d.)
	848 bp	-	
	340 bp	-	
	132 bp	-	
85	910 bp	+	837 bp
	270 bp	-	

were routinely checked for completeness prior to linkering. The smaller HindIII-inserts showed only a small increase in mobility after AluI digestion and were presumed, therefore, to be derived from unique AluI fragments (HindIII linker molecules contain AluI sites; therefore AluI digestion removes short terminal segments from linkered fragments). Only two phage, λIG46 and λIG74 carried HindIII-inserts approximating in size to 194 bp (184 bp T7 DNA and 10 bp linker DNA). The tec1 fragment should yield fragments of 121 bp and 73 bp when digested with AluI together with HaeIII (the 184 bp AluI fragment has a single HaeIII site 116 bp from its left end). However, double-digestion of λIG46 and λIG74 with AluI/HaeIII revealed that neither phage carried a HindIII-insert with an internal HaeIII site (data not shown). These phage, therefore, did not contain inserts carrying T7 tec1.

3.4 Identification of a putative, unknown T7 transcriptional terminator

It had been noted that the phage λIG5 carried an insert of 192 bp which showed only a small increase in mobility after AluI digestion (Fig. 3.3) and contained a HaeIII site (data not shown). Although this phage had not hybridised significantly to the KpnI D probe, it was further examined in case it might prove to carry tec1. Its DNA was digested with HindIII, PstI, BamHI and EcoRI (the latter three enzymes to "smash" the λ vector and so minimise phage-infection problems arising on transformation). The products were then ligated to HindIII-digested pHR1, treated with Calf Intestinal Phosphatase. Following transformation into MC1000, plating on lactose-MacConkey agar containing ampicillin, and overnight incubation at 37°C, colonies were examined for a reduction in lacZ expression. On this medium, MC1000 harbouring plasmids expressing lacZ at a significantly

Fig. 3.3 3-10% Polyacrylamide Gel Electrophoresis

Track	DNA	Restriction Digest
1	Linker oligomers	-
2	T7	AluI
3	pBR322	AluI
4	λ590	AluI
5	λIG5	AluI
6	λIG5	HindIII
7	pBR322	AluI
8	λIG5	HindIII
9	λIG5	HaeIII
10	λ590	HaeIII/HindIII
11	λIG5	HaeIII/HindIII
12	pBR322	HaeIII

Note: The sizes of fragments are given in base pairs.

The marker digests are pBR322/AluI and pBR322/HaeIII. The sizes of these fragments are predicted from sequence data (Sutcliffe, 1978). Note that the 659 bp pBR322/AluI fragment migrates as if it were 722 bp. The position of novel fragments in tracks 5 and 11 are designated 'a'.

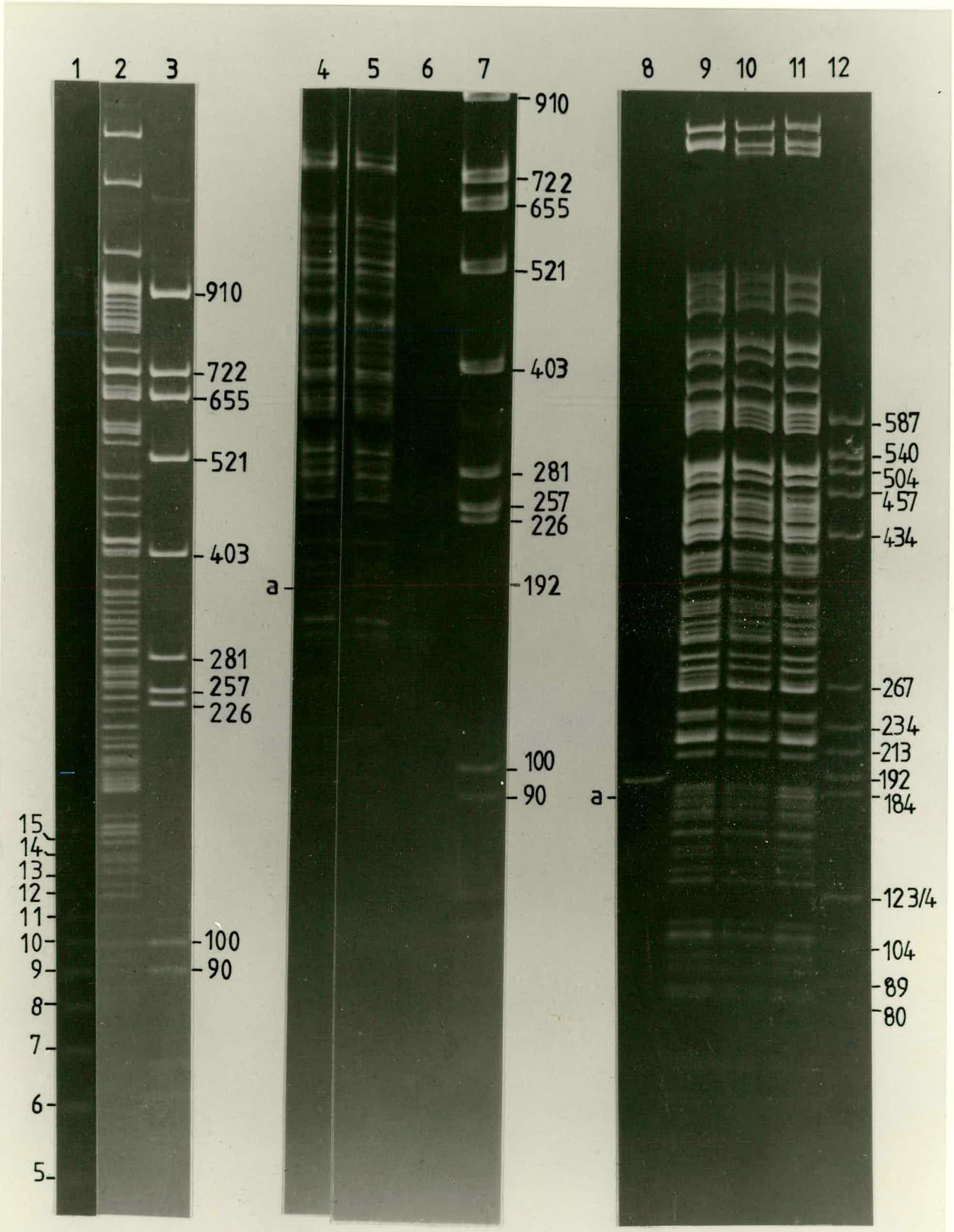

1
2
3
4
5
6
7
8
9
10
11
12
910
722
655
521
403
281
257
226
192
100
90
a
15
14
13
12
11
10
9
8
7
6
5
910
722
655
521
403
281
257
226
192
100
90
a
587
540
504
457
434
267
234
213
192
184
123/4
104
89
80

Fig 3·3

lower level than pHR1 should display a paler phenotype than MC1000 (pHR1) (red). Such a reduction in lacZ expression should be produced by HindIII-inserts carrying transcriptional terminators. This effect was observed: λIG5 yielded 10% white-colony derivatives of pHR1; the remainder were red colonies.

The former were initially characterised by mini plasmid DNA preparations and EcoRI digestion. Fragments of 13.4 Kbp, 7.9 Kbp and 6.1 Kbp are produced by EcoRI digestion of pHR1. Comparison with plasmids derived from six different white colonies revealed that all had lost the 6.1 Kbp EcoRI fragment, and gained a 6.3 Kbp fragment (Fig. 3.4). Since λIG5 contains a single HindIII-insert of 192 bp, and the 6.1 Kbp EcoRI fragment of pHR1 contains its HindIII site, it was suspected that the white colonies contained pHR1 derivatives carrying the 192 bp HindIII-insert from λIG5. This was confirmed by further restriction analysis of CsCl-plasmid DNA preparations using HindIII, each derivative yielding a novel 192 bp fragment.

The position of the HaeIII site within the HindIII-insert was investigated by restriction analysis of both λIG5 and pHR1 derived DNA (Figs. 3.3, 3.4). Double-digestion of λIG5 DNA with HaeIII/HindIII generated a fragment of 180 bp which was absent from the equivalent digest of λNM590. The smaller 12 bp segment of the λIG5 insert was undetectable on the gels used for analysis (3-10% polyacrylamide). Similarly, sequence data predict that pHR1 will yield a 900 bp fragment on digestion with HaeIII, which should be subdivided by HindIII digestion into a 623 bp fragment and a 267 bp fragment (Post et al, 1979; Crawford et al, 1980). These predictions were verified by digestion of pHR1. Digestion of the plasmids derived from the six white colonies with HaeIII revealed, , however,

Fig. 3.4 1% Agarose gel electrophoresis (tracks 1-9) and 3-10% polyacrylamide gel electrophoresis (tracks 10 and 11)

Track	DNA	Restriction digest
1	λcI857	EcoRI
2	p(i)	EcoRI
3	p(ii)	EcoRI
4	p(iii)	EcoRI
5	pHR1	EcoRI
6	p(iv)	EcoRI
7	p(v)	EcoRI
8	p(vi)	EcoRI
9	λcI857	HindIII
10	pHR1	HaeIII
11	pIG103	HaeIII

Note: The sizes of fragments are given in base pairs. The marker tracks (1, 5, 9 and 10) are λcI857 or pHR1. Tracks 2-8 contain "mini" plasmid-DNA preparations, hence the smeary background of E.coli chromosomal DNA. The six plasmids referred to on page 61 are designated p(i)-(vi).

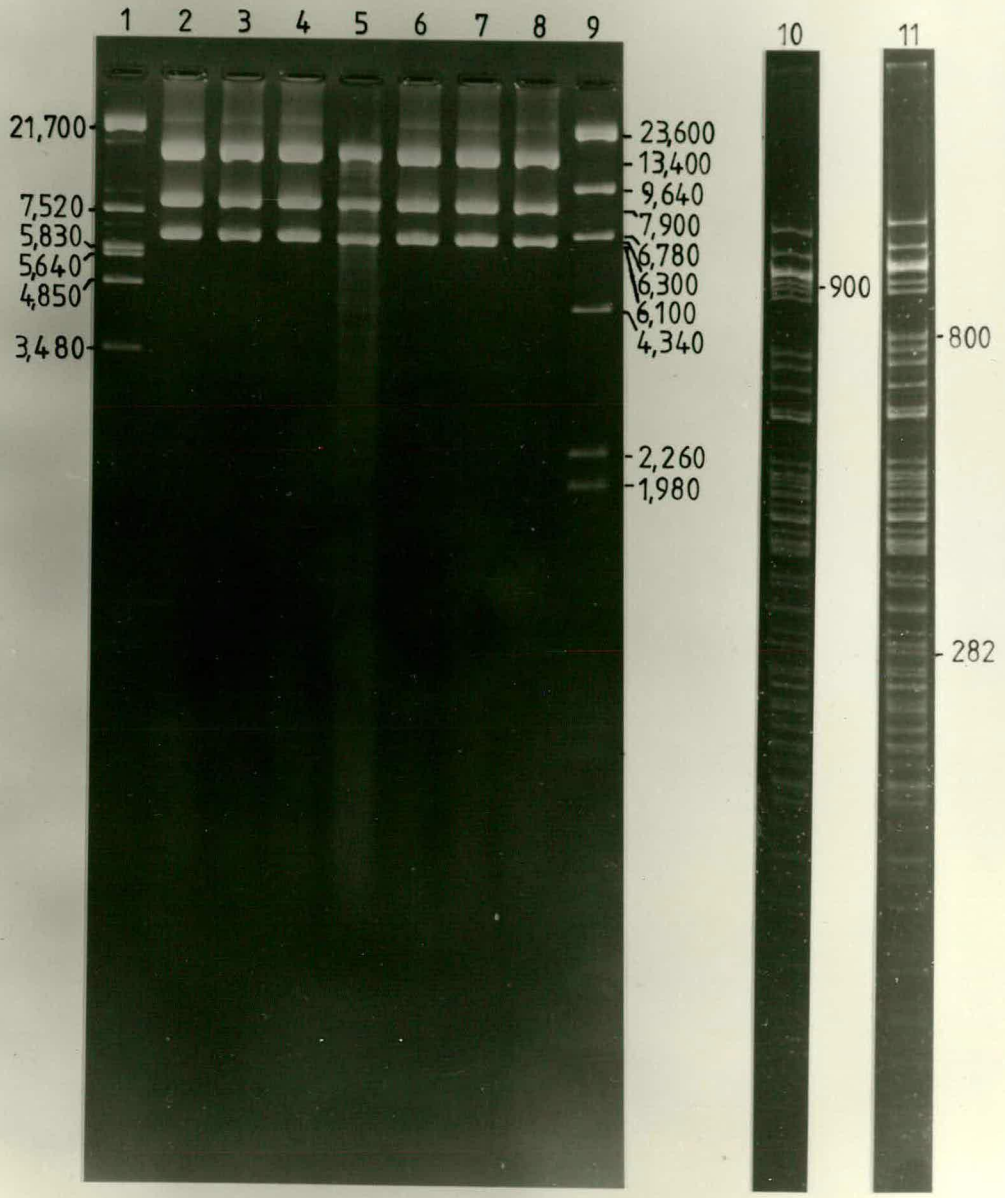

1
2
3
4
5
6
7
8
9
10
11
21,700
7,520
5,830
5,640
4,850
3,480
23,600
13,400
9,640
7,900
6,780
6,300
6,100
4,340
2,260
1,980
900
800
282

Fig 3·4

that the 900 bp fragment was no longer present. Instead, each derivative yielded two extra fragments whose sizes were approximately 800 bp and 282 bp.

These observations suggest: (i) that the six white colonies contained derivatives of pHR1 carrying the 192 bp HindIII-insert from λIG5; (ii) that all six derivatives carried the insert in the same orientation; (iii) that the insert has a HaeIII site approximately 180 bp from one end; (iv) that the insert reduces lacZ expression sufficiently to cause a colour-phenotype change in one orientation only (Fig. 3.4); (v) that the insert did not carry tec1. One of these plasmids was designated pIG103.

3.5 Expression of lacZ by pIG103

To confirm that a reduction in lacZ expression is produced by the 192 bp T7-DNA insert, samples of MC1000 strains harbouring pHR1 and pIG103 were taken between OD_{600} 0.2 and 0.6 and assayed for β-galactosidase activity (Table 3.2). Each result is the mean of 3 or more independent determinations on separate cultures.

The results indicate that MC1000 (pIG103) does indeed exhibit a lower level of β-galactosidase activity than MC1000 (pHR1). It must be concluded that the presence of the 192 bp insert between P_{L10}/P_{L11} and lacZ results in a 78% reduction of lacZ expression. It is most unlikely that this is due to plasmid copy number effects, as plasmid DNA preparations revealed no major differences in yield. However, caution is still needed when interpreting these results as accurate assessment of plasmid copy number is difficult, and artifactual polar effects of fusions can have a strong influence on lacZ expression by pMC81. The results, therefore, merely suggest the possibility that the cutdown of lacZ expression is due to the insertion of a transcriptional terminator in pHR1.

TABLE 3.2

Expression of β-galactosidase by MC1000 (pIG103). The activities of β-galactosidase in MC1000 (pHR1) and MC1000 (pIG103) (expressed in Miller units) are indicated at various stages of growth.

TABLE 3.2

A_{600}	Miller units in MC1000 (pHR1)	Miller units in MC1000 (pIG103)	% Termination
0.2	783.2 (±10)	172.5 (±6)	77.9
0.3	760.7 (±12)	151.8 (±8)	80.0
0.4	755.6 (±16)	166.2 (±10)	78.0
0.5	741.8 (±15)	164.7 (±6)	77.8
0.6	765.1 (±10)	179.4 (±7)	76.6

3.6 Suitable fragments carrying tec1 and tec2

During the course of this work, restriction fragments carrying tec1 and a second transcriptional terminator from T7, tec2 (Minkley and Pribnow, 1973; Peters and Hayward, 1974a), became available from another group (Studier and Rosenberg, 1981). These fragments had been isolated by a similar approach to that described above, but using BamHI synthetic linker molecules. Two fragments containing tec1 had been isolated. The first of these was a 446 bp HpaI fragment (T7 co-ordinates 18.27% to 19.40%) cloned in the BamHI site of pBR322 to produce pAR48, using the following linkers: dCCGGATCCGG. The second was a 142 bp Fnu4HI fragment (T7 co-ordinates 18.77% to 19.12%) cloned in the BamHI site of pBR322 to produce pAR1085, using the following linkers: dGGATCCAGATCCGG (after "filling-in" of the cohesive ends of the Fnu4HI fragment). Finally, the tec2 transcriptional terminator has been isolated on a 632 bp AvaII-HaeIII fragment (T7 co-ordinates 26.70% to 28.28%) cloned in pBR322 to produce pAR410, using the following linkers: dCCGGATCCGG: (AvaII cohesive end "filled-in"). It should, however, be noted that the existence of tec2 in this region has not been demonstrated but is strongly implicated from sequence homology between this fragment and tec1 and current information concerning the position of tec2 in the T7 genome. These plasmids were kindly donated by Dr F. W. Studier, and prevented any further attempts on my part to isolate tec1. However, their inserts could not be studied by subcloning into pMC81, as appropriate cloning sites are not available in this plasmid. The cloning of these fragments into a suitable vector (pHR9) will be described in Chapter 4.

3.7 Discussion

It was disconcerting to find that many of the λ590 derivatives isolated contained more than one inserted HindIII fragment, and that some of these inserts included more than one AluI fragment of T7 DNA. The reasons for this are probably two fold: an insufficient concentration of HindIII linker molecules in the first (blunt- ended) ligation reaction; and an insufficient concentration of HindIII-cut vector DNA in the final ligation. A factor which must have aggravated the first problem is revealed by recent information concerning the nucleotide sequence of the T7 genome (Dunn and Studier, 1982). I initially estimated, by extrapolation from the known AluI map to the left end of T7 DNA, that digestion of the whole genome with AluI should generate some 60 fragments. The complete nucleotide sequence, however, indicates that there are at least 90 fragments. Furthermore, the phosphorylation of the linker molecules may have been inefficient, reducing still further the efficient ratio of linker to T7/AluI fragments. This would favour ligation of T7 AluI to AluI fragments. Poor phosphorylation of linker was in fact indicated, in some preliminary cloning attempts, by an absence of linker oligomers among the products of ligation. The cloning attempt which gave rise to the recombinants described here appeared to be satisfactory, by the latter criterion. However, it might be wiser to monitor the phosphorylation reaction, using γ-^{32}P-ATP, to ensure a high yield and/or known concentration of functional linker molecules.

Underestimation of the number of fragments produced upon digestion of T7 DNA with AluI must also have reduced my chances of recovering a tec1 recombinant from the 93 derivatives of λ590 actually isolated. Owing to the availability of suitable fragments

cloned by other workers, further cloning attempts guided by the above experience were judged to be redundant.

The observed reduction in expression of lacZ by pIG103, as compared with pHR1, must be interpreted with caution. Polarity effects beyond the site of insertion in trpB can certainly influence lacZ expression strongly in this system. The structure of pMC81 (and therefore of pHR1) is such that lacZ is fused in phase with trpA (C.L.Squires, pers. comm.). Efficient expression of lacZ therefore requires efficient initiation of trpA translation. However, the latter is strongly dependent on efficient translation of trpB as trpB and trpA are translationally coupled (c.f. trpE and trpD: Oppenheim and Yanofsky, 1980). In pHR1, the same situation should arise: in this case ribosomes should read from rplJ into trpB in phase (producing a hybrid protein). Insertion of a DNA fragment disrupting trpB translation will therefore have a marked effect on lacZ expression. The likely maximum degree of this effect has been demonstrated by the insertion of a 567 bp 16S rDNA fragment (containing nonsense codons in every reading frame) at the trpB HindIII site in pMC81 (Barry et al, 1979). The result was a 10-fold decrease in arabinose-induced lacZ expression. The insertion of the tec3 fragment at this site in pHR1 might have similar translational effects, thus reducing lacZ expression even if the insert contains no transcriptional terminator. We could only exclude this possibility if the nucleotide sequence of tec3 were known and we could predict that translation of trpB would not be disrupted in pIG103.

Similarly, other possible mechanisms not involving transcriptional termination within the insert cannot be eliminated without further work. For example, the inserted fragment may contain an RNA

processing site, generating 'T7-' trpB-trpA:lacZ mRNA lacking the 5'-end of trpB. This might preclude lacZ translation and/or lead to accelerated degradation of the message. Alternatively, the insert might contain a promoter allowing initiation of transcription "backwards" through trpB into rplJ. Convergent transcription would disrupt expression of lacZ in this instance.

The observation that the drug rifampicin stimulates lacZ expression by pIG103, a phenomenon otherwise observed only with fusions (to lacZ or galK) containing known transcription terminators, suggests that the insert in pIG103 probably does lead to transcriptional termination (Howe et al, 1982; Newman et al, 1982). However, this could either be a secondary result of premature translational termination, leading to exposure of a normally cryptic transcriptional terminator downstream of the insert (recall the proposed mechanism for the polar effects of nonsense mutations: Adhya and Gottesman, 1978); or could reflect the presence of a true transcriptional terminator within the insert. To clarify these questions, I wished to study the tec3 fragment in a fusion system less liable to polarity effects downstream of the cloning site. In the next chapter, I shall describe the insertion and study of tec1, tec2 and tec3 fragments in galK fusion plasmids designed by McKenney et al (1981) to minimise polarity effects.

CHAPTER 4 - CONSTRUCTION OF galK FUSION STRAINS

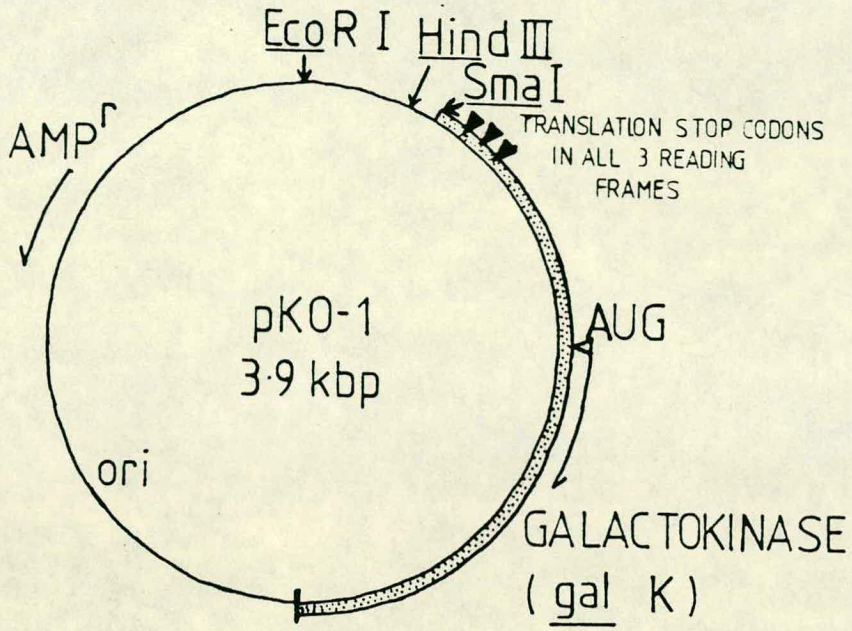
4.1 Introduction

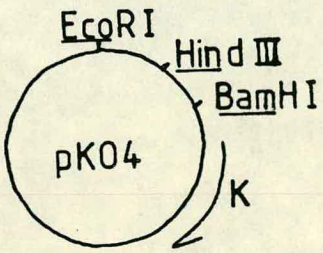
The inherent disadvantages of the pMC81 vector system, and the availability of BamHI restriction fragments carrying T7 tec1 and T7 tec2, prompted the use of a second vector system. The pKO system developed by McKenney et al (1981) combines the concept of gene fusion with recombinant DNA technology to facilitate precise study of transcriptional signals. The approach involves defined fusion of DNA fragments in vitro to plasmid and λ-phage vectors carrying the galactokinase gene (galK). The influence of cloned fragments on galK expression allows identification of those carrying promoter or terminator sequences and determination of their signal strengths. These can then be studied in a multi-copy mode, using plasmid vectors, or a single-copy mode, using λ-phage vectors.

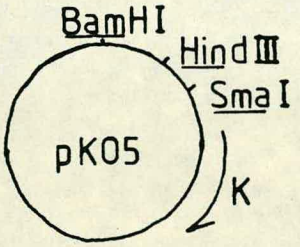
The basic plasmid of the system is a promoter-monitoring vector, pKO1 (Fig. 4.1). This pBR322 derivative possesses the ampicillin resistance gene as an independent selectable marker; the galK coding sequence; and unique restriction sites for the enzymes EcoRI, HindIII and SmaI upstream of galK. The entire nucleotide sequence is known apart from approximately 700 bp internal to the galK gene. This allows the precise cloning of suitable DNA fragments containing known or potential promoter sequences upstream of galK. Derivatives can be screened for promoter activity by colour-phenotype on indicator plates. For example, a galK$^-$ strain carrying pKO1 is white when grown on MacConkey-galactose plates, and red when grown on Tetrazolium-galactose plates because galK is barely expressed. The same strain carrying derivatives of pKO1 containing a correctly orientated promoter-bearing fragment expresses galK more strongly,

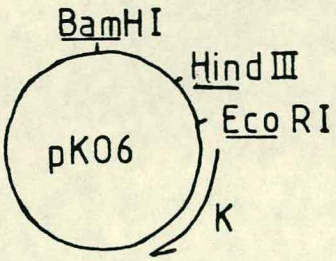
Fig. 4.1 Promoter monitoring plasmids of the pKO system.

The basic plasid, pKO1, is shown together with a number of its derivatives (McKenney et al, 1981).


EcoR I
Hind III
Sma I
AMP^r
TRANSLATION STOP CODONS
IN ALL 3 READING
FRAMES
pKO-1
3·9 kbp
AUG
ori
GALACTOKINASE
(gal K)


EcoRI
Hind III
BamHI
pKO4
K


BamHI
Hind III
Sma I
pKO5
K


BamHI
Hind III
Eco RI
pKO6
K

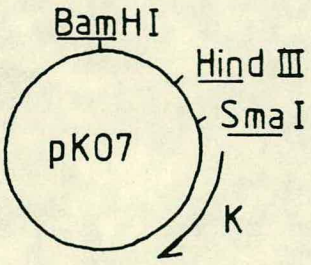

BamHI
Hind III
Sma I
pKO7
K

Fig 4·1

and therefore tends to red on MacConkey-galactose plates and white on Tetrazolium-galactose plates. Furthermore, the efficiency of galK expression can be more accurately determined by assaying the level of galactokinase activity in each strain produced. The assay used is relatively simple, sensitive and linear, detecting between 5 and 200,000 molecules of enzyme per cell.

A further benefit of the choice of galK as monitor gene is its comparative insensitivity to translational polarity. Translation of the galK gene normally initiates independently in an intercistronic region upstream of galK. McKenney et al reasoned that if this region were left intact in the vector, then galK translation might be relatively independent of the RNA structure occurring upstream of the galK gene. Results from experiments in which galK was genetically fused at different distances from the P_L promoter of phage-λ support this contention by demonstrating that galK expression remains proportional to the level of transcription initiated from P_L (Adhya et al, 1976). Accordingly, McKenney et al included gal DNA upstream of galK (168 bp) when constructing the vector.

As a further precaution, they ensured the presence of translation stop codons in all three phases between the cloning sites and the galK AUG. Translation initiating within or upstream of an insert could, in principle, exert differential effects on galK translational efficiency. The above three stop codons are located at least 72 bp upstream of the galK coding sequence, ensuring that there is always a lengthy untranslated leader region.

The properties of the galK gene also permit selection for either positive or negative effects of inserts on gene expression. Activation of the plasmid-encoded galK gene by cloned fragments can

be recognised by restoration of a Gal^+ phenotype to a galK mutant host. Thus, selection by growth on media containing galactose as sole carbon source, or screening on suitable indicator media allows isolation of promoters, promoter "up" mutations and terminator "down" mutations. Similarly, expression of galK in a cell defective in all three gal operon gene functions ($E^-T^-K^-$) leads to the accumulation of the toxic intermediate galactose-1-phosphate, and cell death. This can be exploited to isolate terminators, terminator "up" mutations, and promoter "down" mutations.

However, the plasmid vector system has several limitations, most of which are due to copy number effects. For example: if the transcription signal is affected by ancillary factors, these may become limiting in the cell; the cell may respond abnormally to the presence of many copies of the regulatory site; and if plasmid copy number varies for different constructions, then it is not possible to make valid comparisons of promoter or terminator strengths between them. These problems can be alleviated by transferring the regulatory site-galK fusion to a single gene copy. The complete pKO system allows for this by permitting the transfer of plasmid fusions to a single genomic copy, via a λ-phage intermediate. Despite the potential benefits of the phage system, the present study utilises exclusively plasmid vectors as the requisite phages were unavailable until very recently. (It is not clear that high copy number has, in fact, generated any problems in the present case.)

The vectors used here are all derivatives of pKO1, and were constructed in our laboratory by K. Howe. The unique restriction sites upstream of galK in pKO1 have been manipulated by McKenney and others to generate several different promoter-monitoring vectors (Fig. 4.1). McKenney et al have also constructed a terminator-

Fig. 4.2. Construction of pHR1800 and pHR9.

The EcoRI/HindIII fragment carrying Pgal was derived from pKG1800. This was cloned in pKO1 and pKO4 by K. Howe to produce pHR1800 and pHR9 respectively (see section 4.1). The sizes of these plasmids are shown in kilobase-pairs.

Note: The pKO3 vector fragment of pKG1800 was restricted with PstI to minimise recovery of pKO3 derivatives. Any pKO3 molecules surviving this treatment were detected by HaeII digestion as pKO1 has one less HaeII site than pKO3. This difference arises from the construction of pKO1 from pKO3 by the deletion of 2 bp (at a HaeII site) to produce a vector with all three translation frames closed well upstream of galK.

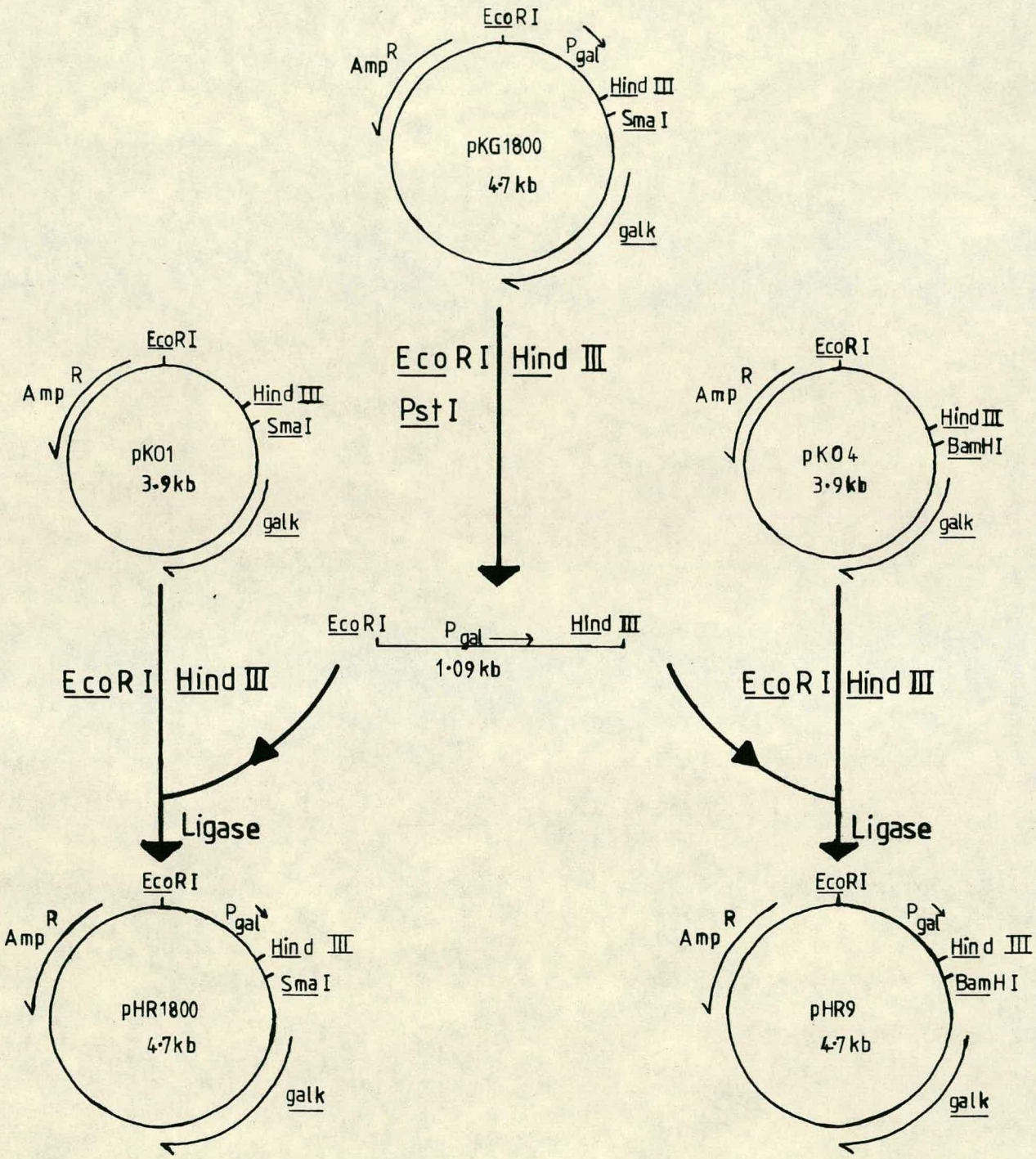

EcoRI
Amp R
Pgal
Hind III
Sma I
pKG1800
4·7 kb
galk
EcoRI Hind III
Pst I
EcoRI
Amp R
Hind III
Sma I
pKO1
3·9 kb
galk
EcoRI
Amp R
Hind III
BamHI
pKO4
3·9 kb
galk
EcoRI Pgal Hind III
1·09 kb
EcoRI Hind III
Ligase
EcoRI Hind III
Ligase
EcoRI
Amp R
Pgal
Hind III
Sma I
pHR1800
4·7 kb
galk
EcoRI
Amp R
Pgal
Hind III
BamHI
pHR9
4·7 kb
galk

Fig. 4·2

monitoring plasmid, pKG1800, in which galK expression is dependent on transcription initiated at the Pgal promoter. However, this vector differs from pKO1 in that the stop codon in one translation phase is separated from the AUG start codon of galK by only 3 bp. It is likely that upstream translation terminating at this stop codon would strongly affect initiation of galK translation. For this reason, K. Howe, in our laboratory, constructed vectors analogous to pKG1800 but derived from pKO1 (Fig. 4.2). Translation of the galK leader region would, therefore, terminate at least 72 bp upstream of the galK coding sequence. The first of these vectors, pHR1800, is derived from pKO1 and pKG1800. The 1.09 kb EcoRI/HindIII fragment carrying Pgal was subcloned from pKG1800 into pKO1, replacing the latter's 0.19 kb EcoRI/HindIII fragment. pHR1800 carries unique restriction sites for HindIII and SmaI between Pgal and galK at which restriction fragments produced by these enzymes (or any enzymes generating flush-ends) can be cloned. The second vector, pHR9, is derived from pKG1800 and pKO4, a promoter-monitoring vector analogous to pKO1. pKO4 had been derived from pKO1 by the insertion of a BamHI linker at its SmaI site. K. Howe transferred the Pgal-containing EcoRI/HindIII fragment from pKG1800 into pKO4 to produce pHR9, a terminator-monitoring vector suitable for studying HindIII and/or BamHI generated fragments.

The use of the above vectors should alleviate problems encountered when using pMC81 and allow the study of the tec1, tec2 and tec3 fragments in a single system. It should, however, be noted that, as discussed in detail later, the study of terminators even in the pKO system is complicated owing to the lack of control over translation in the region immediately upstream of the transcriptional terminator. I shall describe the construction of the

desired fusions in numerical rather than historical order (i.e. tec1, tec2, tec3).

4.2 Construction of fusions carrying tec1 fragments

Initial attempts to subclone the BamHI-inserts from pAR48 and pAR1085 into pHR9 were unsuccessful. Donor DNA was digested with BamHI, EcoRI, HindIII and PstI (the latter three enzymes to "smash" the pBR322 vector) and ligated to BamHI-digested pHR9 DNA treated with calf intestinal phosphatase. Following transformation into N100, plating on galactose-MacConkey agar containing ampicillin, and overnight incubation at 37°C, the resulting colonies were either red or white. The red colonies had the same phenotype as N100 (pHR9). They could well have been re-created N100(pHR9), but might also have included N100 harbouring pHR9/tec1 recombinants if these produced enough galactokinase to develop the red colony phenotype. The white colonies could similarly have represented N100 transformed with two kinds of plasmids: pBR322 (or pBR322/tec1); or pHR9/tec1 recombinants if these produced insufficient galactokinase to allow the red colour to develop. Restriction analysis of mini plasmid DNA preparations from 20 red and 20 white colonies derived from each ligation suggested that the red colonies were all N100(pHR9) and the white colonies were N100(pBR322), N100(pAR48) and N100(pAR1085) (data not shown). This conclusion was supported by the observation that some white derivatives were also tetracycline resistant.

Two other subcloning strategies, aimed at minimising recovery of the pBR322 vector, were considered. The tec1 fragments could either be purified and used in ligations to pHR9, or an intermediate vector could be used to allow selection against donor vector molecules. Since further subclonings might be necessary in future, I reasoned that the second strategy was more attractive. I therefore subcloned

each tec1 fragment into the plasmid pACYC184. This is a 3.97 kb plasmid conferring chloramphenicol and tetracycline resistance on its hosts (Chang and Cohen, 1978). Within this plasmid, insertions at a unique EcoRI restriction site disrupt the chloramphenicol resistance gene, while insertions at the unique HindIII, BamHI or SalI restriction sites disrupt the tetracycline gene. pAR48 and pAR1085 DNA were digested with BamHI and ligated separately to BamHI-digested pACYC184. Following transformation into ED8654, plating on L-agar containing chloramphenicol, and incubation overnight at 37°C, colonies were transferred, by cocktail stick, in an ordered array to L-agar containing chloramphenicol, tetracycline or ampicillin. After overnight incubation at 37°C, 10% of picked colonies exhibited a Chl^R Tet^S Amp^S phenotype, as expected for ED8654 strains harbouring pACYC184 derivatives containing, at their BamHI sites, inserts which do not include the amp^R DNA of pBR322.

The tetracycline resistance gene of pACYC184 is derived from the same source as that in pBR322, and should, therefore, possess the same nucleotide sequence. Accordingly, pACYC184 should carry a 104 bp HaeIII fragment which can be subdivided with BamHI into 80 bp and 24 bp (Sutcliffe, 1978). These predictions were verified by digestion of pACYC184. The physical structure of plasmid DNA from Chl^R Tet^S Amp^S colonies derived from both ligations was then investigated using mini plasmid DNA preparations. Digestion with EcoRI, to linearise each plasmid, revealed that, as expected, the putative pACYC184/pAR48 recombinants were about 4.6 kb, and the putative pACYC184/pAR1085 recombinants were about 4.15 kb. Digestion with HaeIII revealed that they had all lost the 104 bp fragment and gained new fragments of the expected sizes (data not shown). One plasmid derived from the pACYC184/pAR48 ligation was designated

pIG124, and one derived from the pACYC184/pAR1085 ligation was named pIG132. ED8654 strains harbouring these plasmids were used to prepare plasmid DNA by CsCl-gradient-centrifugation.

pIG124 DNA was digested with BamHI and ligated with BamHI-digested pHR9 DNA treated with calf intestinal phosphatase. Following transformation into N100, plating on galactose-MacConkey agar containing ampicillin, and overnight incubation at 37°C, only red colonies were observed (the same phenotype as N100(pHR9)). 100 colonies were tested for chloramphenicol resistance by picking to L-agar containing this antibiotic. It was noted that 6 colonies were "sickly" on galactose-MacConkey agar containing ampicillin. These alone proved to be Chl^R, and could have arisen by double transformation with pHR9 and pACYC184 or by transformation with a pHR9/pACYC184 hybrid. The colonies were screened further by hybridisation with a [^{32}P]-T7 DNA probe. 24 colonies responded positively. 10 of these were analysed by restriction of mini plasmid DNA preparations. These results suggested that all were pHR9 derivatives carrying the 456 bp BamHI tec1 fragment. Two restriction patterns were observed, suggesting that the fragment had been cloned in both orientations. The plasmid containing the fragment in the 'e'-orientation (transcribed in the same direction as in T7) was designated pIG125. The plasmid containing the fragment in the 's'-orientation (inverted) was designated pIG126. Restriction analysis of CsCl-gradient plasmid DNA preparations from N100 strains harbouring these plasmids confirmed that their structures were as expected (Fig. 4.3, Table 4.1 and Fig. 4.4). Note that both plasmids have lost the 930 bp, HaeIII fragment of pHR9, which should be subdivided by BamHI digestion to yield fragments of 651 bp and 279 bp (running as 650 bp and 280 bp in my gels). Note also that

Fig. 4.3

Proposed structures of the constructed galK fusions.

Restriction sites are represented as follows: HaeIII, ⚲ ; HindIII, ⍎ ; BamHI, ↑ . The distances between restriction sites are given in base-pairs. The sizes of the 190 bp HindIII/HaeIII fragment of pHR1800 and the 279 bp BamHI/HaeIII fragment of pHR9 are predicted from sequence data (McKenney et al, 1981). Similarly, the sizes of T7 DNA fragments expected from pIG125, pIG126, pIG134, pIG135, pIG138 and pIG139 are also predicted by sequence data (Dunn and Studier, 1981). The sizes of the 630 bp HaeIII/HindIII fragment of pHR1800, the 651 bp HaeIII/BamHI fragment of pHR9 and the T7 DNA fragments expected from pIG115, pIG116 and pIG117 are derived from restriction analyses reported in the text (sections 4.4 and 3.4). The latter are in good agreement with a more precise determination of these distances by DNA sequencing (Chapter 5).

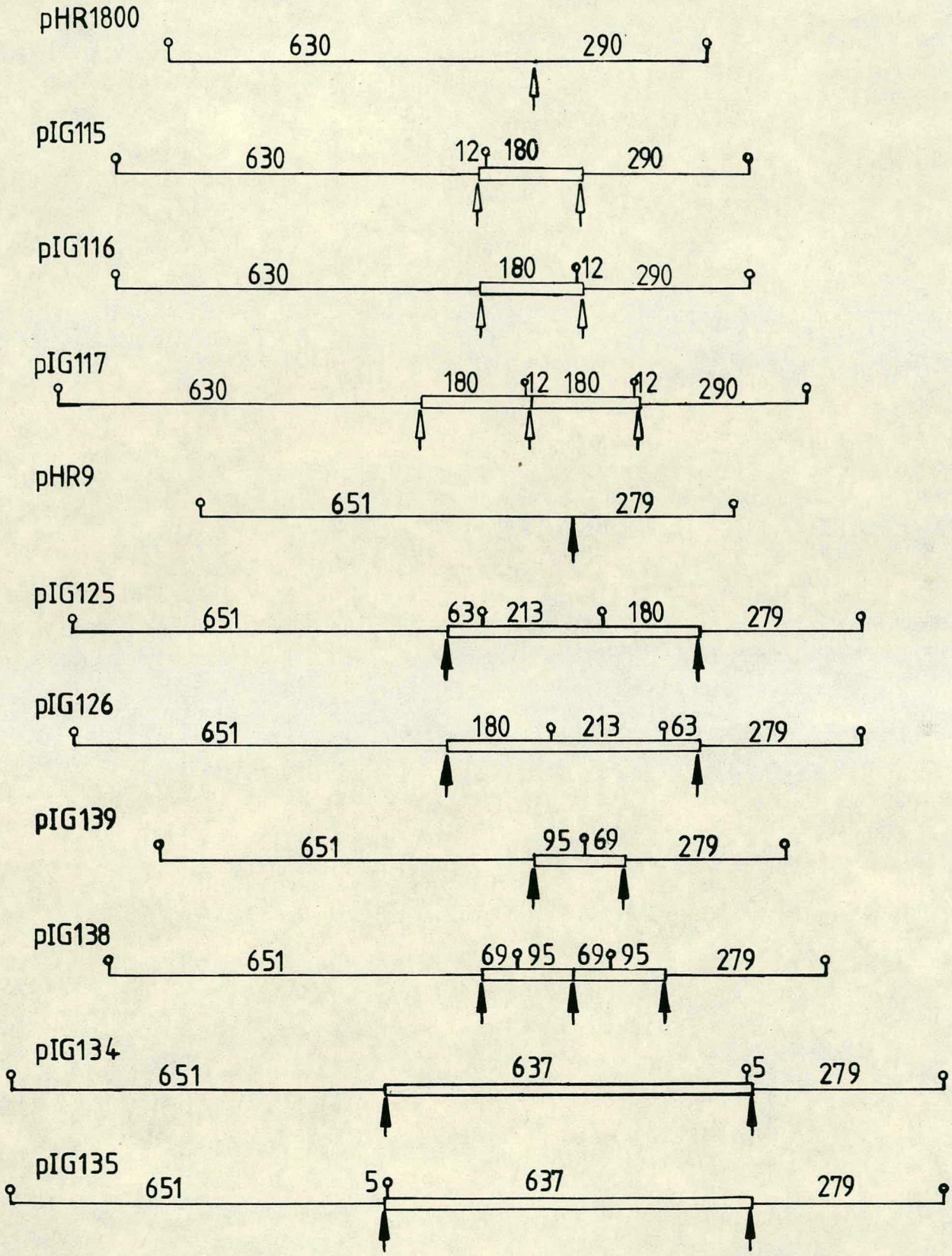

pHR1800
630
290
pIG115
630
12
180
290
pIG116
630
180
12
290
pIG117
630
180
12
180
12
290
pHR9
651
279
pIG125
651
63
213
180
279
pIG126
651
180
213
63
279
pIG139
651
95
69
279
pIG138
651
69
95
69
95
279
pIG134
651
637
5
279
pIG135
651
5
637
279

Fig.4·3

TABLE 4.1

Plasmid	Restriction	Observed Sizes of New Restriction Fragments not present in vector (bp)	Expected Sizes of New Restriction Fragments
pIG115	HaeIII	650 and 450	662 and 470
	HaeIII/HindIII	180	180
pIG116	HaeIII	820 and 305	831 and 302
	HaeIII/HindIII	180	180
pIG117	HaeIII	820, 305 and 190	831, 302 and 192
	HaeIII/HindIII	180	180
pIG125	HaeIII	730, 447 and 215	714, 459 and 213
	HaeIII/BamHI	215, 180 and 65	213, 180 and 63
pIG126	HaeIII	850, 347 and 215	831, 342 and 213
	HaeIII/BamHI	215, 180 and 65	213, 180 and 63
pIG134	HaeIII	1282 and 288	1288 and 284
pIG135	HaeIII	900 and 665	916 and 656
pIG138	HaeIII	706, 370 and 170	720, 375 and 165
pIG139	HaeIII	722 and 344	747 and 348

A summary of the restriction analyses of the galK fusions constructed. The estimated sizes of novel fragments, not observed on digestion of vector DNA, are given. Sequence data for T7 (Dunn and Studier, 1981) and the observed position of the HaeIII site in the 192 bp HindIII-insert from λIG5, allow calculation of the expected sizes of novel fragments for comparison with my actual estimates (see also Fig. 4.3). The 920 bp HaeIII fragment containing the HindIII site of pHR1800 was not present in pIG115, pIG116 or pIG117. Similarly, the 930 bp HaeIII fragment containing the BamHI site of pHR9 was not present in pIG125, pIG126, pIG135, pIG138 and pIG139.

Fig. 4.4 3-10% Polyacrylamide Gel

Track	DNA	Restriction Digest
1	pBR322	HaeIII
2	pIG125	BamHI
3	pHR9	HaeIII/BamHI
4	pHR9	HaeIII
5	pIG126	HaeIII
6	pIG125	HaeIII
7	pHR9	HaeIII/BamHI
8	pIG126	HaeIII/BamHI
9	pIG125	HaeIII/BamHI
10	pBR322	HaeIII

Note: Sizes of fragments are given in base pairs.

The marker tracks are pBR322/HaeIII. The 930 bp/HaeIII fragment of pHR9 is shown intact, and subdivided by BamHI digestion to yield fragments of 651 bp and 279 bp. The sizes of novel fragments in tracks 5, 6, 8 and 9 are shown.

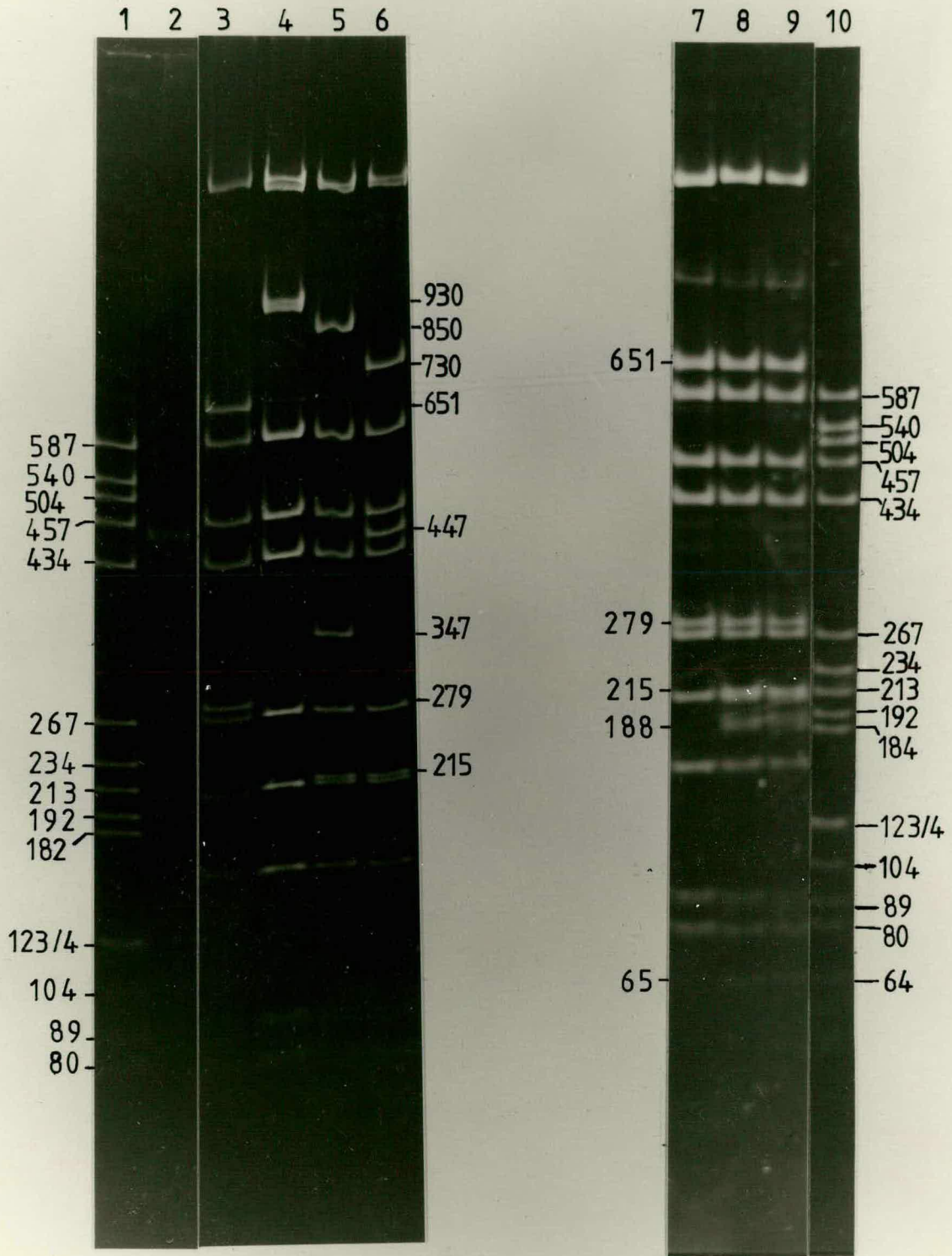

1 2 3 4 5 6
7 8 9 10
930
850
730
651
587
540
504
457
434
447
347
279
267
234
213
192
182
215
123/4
104
89
80
651
279
215
188
587
540
504
457
434
267
234
213
192
184
123/4
104
89
80
65
64

Fig 4·4

the phenotypes of N100(pIG125) and N100(pIG126) were in the reverse order to that expected in simple terms; reduction of galK levels had been hoped for in the former but not the latter strain. However, cutdown was detected by galK assays in both cases (Table 4.2), but only the pIG126 derivative showed a noticeably altered colour phenotype.

pIG132 DNA was treated similarly, again giving rise only to red colonies; 5 of these were "sickly" and Chl^R. Hybridisation with [^{32}P]-T7 DNA revealed that 32% of these colonies contained DNA homologous to the probe. Restriction analysis of mini plasmid DNA preparations from 10 of these suggested that they were pHR9 derivatives carrying the 164 bp BamHI tec1 fragment. The two restriction patterns observed suggested that this fragment had been cloned both in the 'e'-orientation and as two tandem copies in the 's'-orientation. A plasmid containing a single-copy of this fragment in the 's'-orientation has not yet been recovered. The corresponding plasmids were designated pIG139 and pIG138 respectively. Restriction analysis of CsCl-gradient plasmid DNA preparations from N100 strains harbouring pIG138 and pIG139 confirmed that their structures were as expected (Fig. 4.3, Table 4.1 and Fig. 4.5). Note the interruption of the 930 bp HaeIII fragment of pHR9.

4.3 Construction of fusions carrying the tec2 fragment

pAR410 DNA was treated similarly, by cloning the insert first into pACYC184 and subsequently into pHR9. Ligation of BamHI-digested pAR410 with BamHI-digested pACYC184 gave rise to 25% Chl^R Tet^S Amp^S. Restriction analysis of mini plasmid DNA preparations from 10 of these suggested that they were pACYC184 derivatives carrying the 642 bp BamHI tec1 fragment. One of these plasmids was designated pIG129 and its DNA was purified by the CsCl-gradient method. It

Fig. 4.5 3-10% Polyacrylamide Gel Electrophoresis

Track	DNA	Restriction Digest
1	pBR322	HaeIII
2	pIG134	BamHI
3	pIG139	BamHI
4	pIG134	HaeIII
5	pIG135	HaeIII
6	pHR9	HaeIII
7	pIG138	HaeIII
8	pIG139	HaeIII

The sizes of fragments are given in base pairs.

The marker track is pBR322/HaeIII. The sizes of novel fragments in tracks 4, 5, 7 and 8 are shown. The 930 bp/HaeIII fragment of pHR9 containing its BamHI site is also shown.

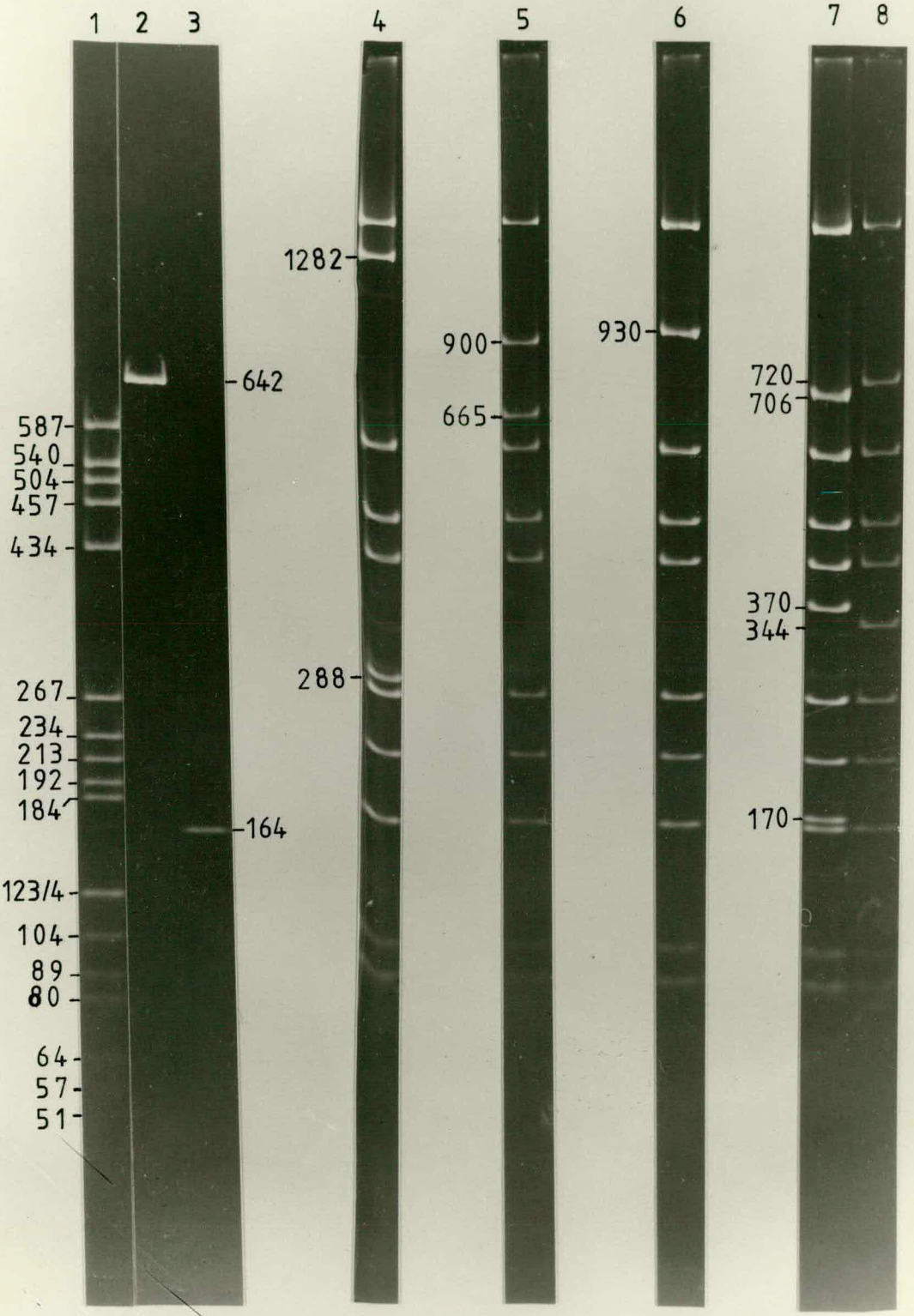

1
2
3
4
5
6
7
8
642
587
540
504
457
434
267
234
213
192
184
164
123/4
104
89
80
64
57
51
1282
288
900
665
930
720
706
370
344
170

Fig 4·5

was then digested with BamHI, and ligated to BamHI-digested pHR9 treated with calf intestinal phosphatase. Transformation into N100 again gave rise only to red colonies. Hybridisation of 100 of these to a [^{32}P]-T7 DNA probe revealed that 22 of them had homology to T7 DNA. Restriction analysis of mini plasmid DNA preparations suggested that they were pHR9 derivatives carrying the 642 bp BamHI tec2 fragment. The two restriction patterns observed indicated that the fragment had been cloned (in single-copy) in both orientations. The plasmid carrying the fragment in the 'e'-orientation was designated pIG134, whereas the plasmid carrying the fragment in the 's'-orientation was named pIG135. Restriction analysis of CsCl-gradient-purified plasmid from strains harbouring these plasmids confirmed their expected structures (Fig. 4.3, Table 4.1 and Fig. 4.5). Note that they lack the 930 bp HaeIII fragment characteristic of pHR9.

4.4 Construction of fusions carrying the tec3 fragment

λIG5 DNA was digested with HindIII and EcoRI (the latter enzyme to "smash" the λ-vector and so prevent transfection during transformation) and ligated to HindIII-digested pHR1800 DNA treated with calf intestinal phosphatase. Transformation gave rise to two colony phenotypes: 97% red colonies and 3% white colonies. This was the first subcloning attempted in the pKO system, and it was suspected that the white colonies would harbour pHR1800-derivatives containing the tec3-insert in the same orientation as in pIG103 (Chapter 3) and therefore yielding less galactokinase than pHR1800. For this reason, these colonies were not screened by hybridisation but directly by restriction analysis of mini plasmid DNA preparations from 4 white colonies and 18 red colonies (the latter to search for tec3 inserted in the inverse orientation). Unexpectedly, the

Fig. 4.6 3-10% Polyacrylamide Gel Electrophoresis

Track	DNA	Restriction Digests
1	pHR1800	AluI
2	pBR322	HaeIII
3	pIG116	AluI
4	pIG116	HindIII
5	pHR1800	HaeIII
6	pBR322	HaeIII
7	pIG117	HaeIII
8	pIG116	HaeIII
9	pIG115	HaeIII
10	pHR1800	HaeIII/HindIII
11	pIG116	HaeIII/HindIII

The sizes of fragments are given in base pairs.

The marker tracks are pBR322/HaeIII. The sizes of novel fragments in tracks 3, 7, 8, 9 and 11 are shown. The 920 bp/HaeIII fragment of pHR1800 is shown intact (track 5) and subdivided with the HindIII to yield two fragments of 630 bp and 290 bo (tracks 10 and 11).

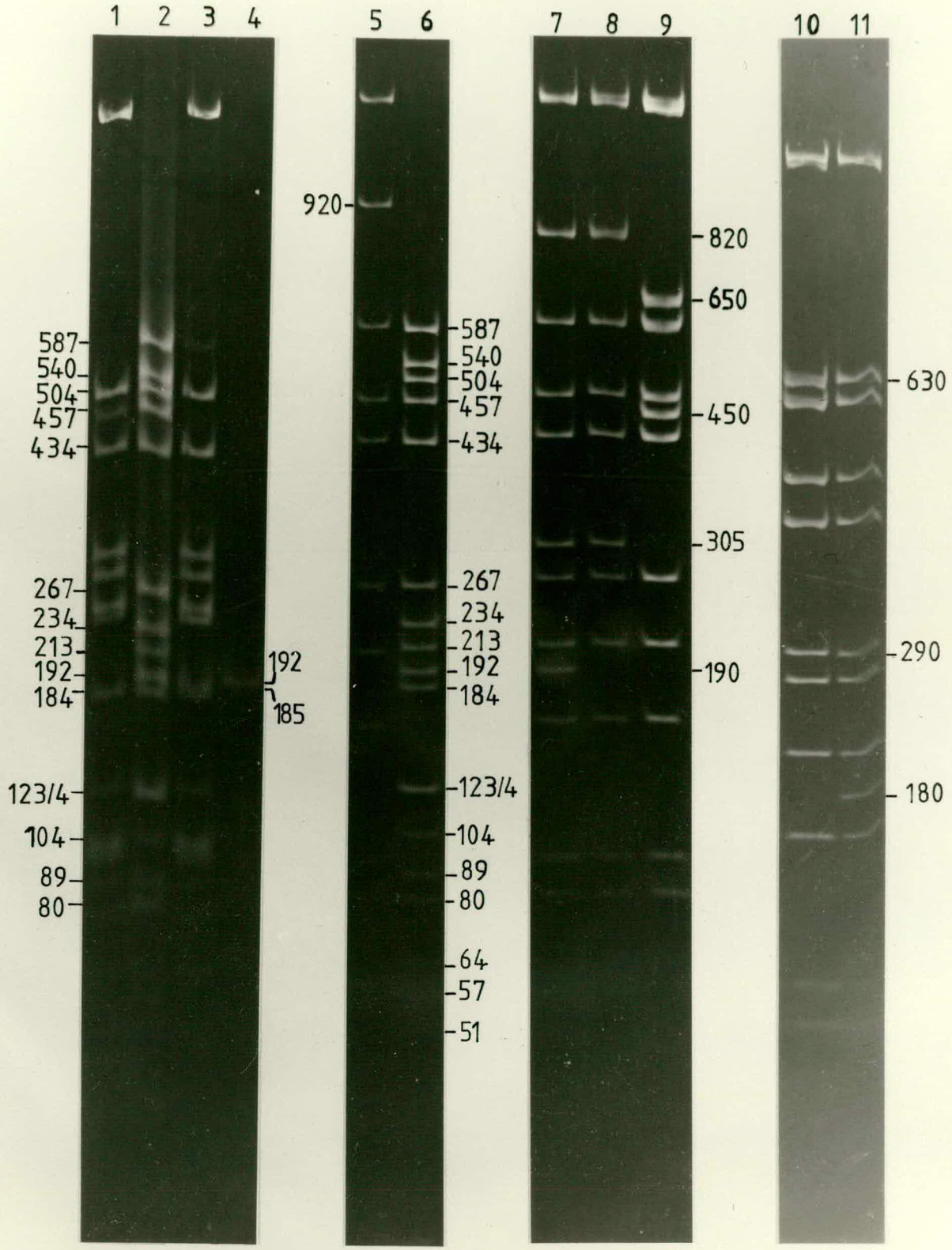

1 2 3 4
5 6
7 8 9
10 11
920
587
540
504
457
434
267
234
213
192
184
123/4
104
89
80
64
57
51
820
650
450
305
190
630
290
180
192
185

Fig 4-6

results indicated that: the 4 white derivatives contained plasmids carrying two tandem copies of the tec3-fragment in the 'e'-orientation (i.e. the same transcriptional orientation as in pIG103); 6 red derivatives contained plasmids with single inserts in either the 'e'- or the 's'-orientation; 1 red derivative contained a pHR1800/λ recombinant; and 9 red derivatives contained pHR1800 itself. It was these findings, together with the results of the galK assays on these strains (Table 4.2) and the phenotype of colonies derived from attempts to subclone tec1 and tec2 fragments, which prompted the introduction of the colony-hybridisation step in the screening procedure for tec1 and tec2 clones (already described).

One plasmid containing the tec3 fragment in the 'e'-orientation was designated pIG116 and one with the 's'-orientation was named pIG115. One plasmid containing two tandem copies in the 'e'-orientation was designated pIG117. Restriction analysis of CsCl-gradient plasmid DNA preparations from strains harbouring these plasmids confirmed their structure (Fig. 4.3, Table 4.1 and Fig. 4.6). Note the absence of the 920 bp HaeIII fragment of pHR1800, which is subdivided by HindIII digestion into 630 bp and 290 bp.

4.5 Expression of galK by the novel fusions

N100 strains harbouring the above fusions were assayed for galactokinase activity. Comparison with galactokinase levels in N100(pHR1800), N100(pHR9) and N100(pK01) revealed that some of these strains did indeed exhibit a reduced expression of galK (Table 4.2). N100(pIG125) shows a 78% reduction in galK levels when compared to N100(pHR9). Similarly, N100(pIG136) shows a 98% reduction in galK levels. The latter strong cutdown was unexpected. *A priori* it might be due to the inverted tec1 terminator *per se*, or the

TABLE 4.2

The results of galK assays on N100 strains harbouring tec1, tec2 and tec3-galK fusions. Each result is derived from assays on three separate cultures. galK units are expressed in nanomoles of galactose phosphorylated per minute per A_{650}. The percentage termination is given where relevant. The colony colour phenotype of each strain on either galactose-MacConkey-ampicillin agar (22-26 hour incubation) or galactose-Tetrazolium-ampicillin agar (14-16 hour incubation) is recorded.

TABLE 4.2

Plasmid	Insert	galK Units	% Termination	Gal-Mac-Amp		Gal-Tet-Amp	
				centre	halo	centre	halo
pHR1800		570 (±21)	-	Red	White	Pink	White
pIG115	tec3 ←	544 (±16)	5	Red	White	Pink	White
pIG116	tec3 →	30 (±4)	95	Pink	Pink	Pale Red	White
pIG117	tec3 → tec3 →	0	100	White	White	Red	Red
pKO1		11 (±1)	-	Pink	White	Red	Pink
pHR9		1029 (±50)	-	Red	White	Pink	White
pIG125	tec1 → 456 bp	226 (±8)	78	Red	White	Pink	White
pIG126	← tec1 456 bp	20 (±3)	98	Pink	Pink	Red	Pink
pIG139	tec1 → 164 bp	243 (±10)	76	Red	White	Pink	White
pIG138	← tec1 164 bp ← tec1 164 bp	178 (±7)	83	Red	White	Pink	White
pIG134	tec2 →	1124 (±60)	0	Red	White	Pink	White
pIG135	← tec2	15 (±2)	99	Pink	Pink	Red	Pink

coincidental features of the inverted 456 bp DNA insert. The galactokinase levels observed in the N100 strains harbouring pIG139 and pIG138 indicate that the latter explanation may be true. N100(pIG139) shows a 76% reduction in galK levels, in good agreement with N100(pIG125). N100(pIG138), however, shows only an 83% reduction in galK levels despite containing two copies of tec1 in the same orientation as in pIG126. This suggests that each 164 bp (inverted tec1) fragment is responsible for an approximate 60% reduction of galK expression in pIG138. Whether this effect is due to the inverted tec1 terminator itself and/or to other sequences in the fragment, the results suggest that the 98% reduction in galK seen in N100(pIG126) is not exclusively due to the inverted tec1 sequence. Other sequences within the 456 bp insert in pIG126, or at its fusion junctions must play a role. In view of the properties of the pKO system, these effects are probably transcriptional rather than translational.

N100(pIG134) shows a somewhat surprising result in that no reduction in galK levels is observed. Two possibilities exist to explain this: (i) transcriptional termination does not occur; (ii) the fusion possesses a promoter of similar strength to Pgal, downstream of the terminator sequence. The latter explanation would allow transcriptional termination to occur, without it being detectable by galactokinase assay. This alternative was easily tested by subcloning the 642 bp BamHI fragment from pIG129 into pKO4. Transformation and plating on galactose-MacConkey agar containing ampicillin gave rise only to white colonies (as N100(pKO4)). Should such a promoter exist on this fragment, recombinants carrying it in the correct orientation would give rise to red colonies (as N100(pHR9)). It therefore seems unlikely that the observed galK

levels in N100(pIG134) are due to the existence of tec2 and a promoter on the 642 bp fragment. N100(pIG135) shows another unexpected result: a 99% reduction in galK levels. Again, this could be due to the inverted (putative) tec2 sequence, or to other sequences within or at the junctions of the insert.

The results obtained with tec3-galK fusions support the hypothesis that the reduction in lacZ expression produced by the same insert in pMC81 (generating pIG103) was due to transcriptional termination at the putative "tec3" site. N100(pIG116), with tec3 in the same orientation as in pIG103, shows a 95% reduction in galK levels. N100(pIG117), with two copies of the fragment in the 'e'-orientation, shows no expression of galK. I infer that the tec3 fragment causes termination of transcription (with 95% efficiency) in this orientation in the pKO system. N100(pIG115), with tec3 in the 's'-orientation, shows only a 5% reduction in galK levels (which is a negligible effect in the standard errors of the results).

It was surprising that the colour phenotypes of the strains isolated were not more distinct from those of strains carrying the vector plasmids. After overnight incubations (i.e. 14-16 hours) on galactose-MacConkey agar containing ampicillin at 37°C, it was impossible to distinguish strains producing as few as 20 galK units from those with up to 1000 galK units. In an attempt to improve the sensitivity of this test, strains were streaked on the same medium and growth observed at regular intervals from 0 to 30 hours. In this way, it was found that distinctions could be made between strains harbouring the vectors (pHR1800 or pHR9; 570 and 1029 units respectively) and those showing 95% termination (@ 30 units). Similarly, strains were grown on galactose-Tetrazolium agar containing ampicillin where the colour phenotype of Gal^+ and Gal^- strains is

reversed (Table 4.2). This screening proved less demanding in terms of incubation times, but again could distinguish between 0 and 95% termination, but not between 0 and 83%. On the MacConkey medium, an incubation of 22-26 hours was required for these distinctions to be made, whereas the Tetrazolium medium required only a 14-16 hour incubation. Furthermore, the distinctions on Tetrazolium plates became more obvious after further incubation at 37°C or 20°C.

The observed difference in galK activity between N100(pHR1800) and N100(pHR9) is probably due to a plasmid copy number effect. N100(pHR9) routinely gives higher plasmid yields than N100/pHR1800 (2-4 fold). This variation is mysterious, as pHR9 differs from pHR1800 by only 10 bp; it is as yet unexplained. However, the derivatives of each vector give DNA yields similar to their parents, so that valid comparisons of galactokinase levels are possible.

Finally, it was surprising to find a reduction in galK levels by several "inverted" terminator fragments. Consideration of the sequences transcribed in each fusion is necessary for interpretation of these results.

4.6 Discussion

It should be noted, that in the following discussion I assume that all observed effects are transcriptional. This is likely in view of the properties of the galK system, but not definite. Other effects, e.g. RNA processing, may be involved and cannot be discounted until the completion of *in vitro* transcription studies on these fusions.

The observed reduction in galactokinase expression caused by the inserts in pIG125 and pIG139 suggests an efficiency of termination for tec1 *in vivo* which is in good agreement with the estimate

of about 75% made, for T7 itself, by McAllister and Barrett (1977). However, this efficiency is somewhat lower than that observed in vitro by Millette et al (1970) and Peters and Hayward (1974), i.e, 90-95%. This discrepancy may reflect an imperfect reconstruction in vitro of the intracellular situation due, for example, to the omission of an as yet undefined cofactor(s) involved in the termination event at tec1. We can, therefore, speculate on the in vivo situation, in the light of the new information that short tec1-bearing fragments, when cloned in the pIG plasmids, mimic the behaviour of tec1 in T7 itself. Assuming that one or more protein factors are missing, they are unlikely to be required for the first steps of termination at tec1, as in their absence in vitro, these first steps occur with an efficiency of at least 95%. However, the results of O'Hare and Hayward (1981) suggest that the termination process is not efficiently completed under these conditions. Thus, it is possible that the factor(s) shift a termination intermediate (formed with 95% efficiency in vitro) into a new conformation, from which 75% of complexes rapidly proceed to complete termination, while 25% proceed to readthrough. The first intermediate referred to might, for example, be a "paused" complex (termination in vitro may follow an alternative non-physiological pathway to completion).

The alternative possibility is that the cofactor(s) stimulate(s) readthrough of tec1. Assuming a protein:protein interaction for antitermination, we have at least four possible ways to explain the observed results: (i) the factor(s) may interact with Rpol shortly after transcriptional initiation, and associate with it during elongation. If the initial "loading" is only 25% efficient then 75% of Rpol molecules will terminate transcription at tec1; (ii) in the same situation, if the initial "loading" were stoichiometric,

a 75%-efficient "unloading" event upstream of tec1 would prevent stoichiometric readthrough of the terminator; (iii) the factor(s) may interact with Rpol as it approaches tec1, affecting only 25% of transcribing Rpol molecules as above; (iv) loading may be stoichiometric but the factor(s) may be only 25% efficient in catalysis of readthrough. In any case, we can predict the occurrence of sequences analogous to the nut sites of λ (Chapter 1) appropriately situated in the T7 genome. Moreover, these features must be retained or mimicked in the pIG plasmids.

Ponta et al (1974) suggested that the 0.7 gene product is required for efficient termination at tec1 in vivo. If this were the case, we would expect less efficient termination in vivo when tec1 is studied in isolation from gene 0.7, i.e. in N100 (pIG125) and N100 (pIG139). The observation that tec1 stimulates 76-78% termination in the pKO system, as in T7 itself, suggest that the 0.7 gene product is not required for efficient termination at tec1. However, we cannot exclude the inelegant alternative hypothesis that T7 infection alters the termination properties of E.coli such that gene 0.7 is needed to restore the status quo (at least in respect of tec1).

Inversion of the tec1 fragments produces somewhat surprising results in that we might expect the inverted tec1 DNA sequence to function much less efficiently as a terminator than tec1 itself. Calculations based on the "rules" of Tinoco et al (1973) indicate that the stem-loop component should have a free energy of -14.4 Kcal in its normal orientation, but only -12.4 Kcal when inverted (due to the presence of a disruptive C-A mismatch in the inverted structure). There is precedent for supposing that this is significant: introduction of C-A mismatches in the stems of trpat(3:4) and λt_{R1} reduces the free

energy of these structures (admittedly more markedly than for tec1) and greatly increases transcriptional readthrough in both cases (Stauffer et al, 1978; Zurawski and Yanofsky, 1980; McDermit et al, 1976). Even supposing that the weakened stem-loop of the inverted tec1 sequence is sufficiently strong to allow termination, termination at this site would further require that translation does not prevent formation of the RNA stem-loop structure, and that appropriate "tail" sequences occur downstream of this structure. Examination of the pIG126 sequence (Fig. 4.8) reveals that translation should terminate 30 nucleotides upstream of the inverted tec1 stem-loop thus allowing it to form. However, the downstream sequence has only four, dispersed T residues out of 11 and also bears no resemblance to known rho-dependent terminator "tails". Hence, it seems unlikely that the tec1 sequence per se could produce strong transcriptional termination in the inverse orientation.

In reality, termination is observed both in N100 (pIG126) and N100 (pIG138). However, the 98% termination observed in the former plasmid cannot be due exclusively to the inverted tec1 sequence, because the shorter inverted tec1-containing fragment present in pIG138 causes only 84% termination, despite being present in tandem copies. (We can estimate that this fragment would cause 60% termination in single copy.) Termination at the inverted tec1 site should be immune to translational interference as translation of both pIG126 and pIG138 should stop 30 nucleotides upstream of the first inverted tec1 stem-loop encountered. Hence, we must consider the possibility that other sequences within (or at the boundary of) the fragment cloned in pIG126 may be causing additional transcriptional termination. Before doing so, a digression to consider the present difficulty of recognising potential terminators in a known sequence is called for.

A nucleotide sequence can be scanned by computer to predict regions of homology capable of forming stem-loop structures. This usually reveals a wide variety of relatively stable structures. Whether or not these should be considered putative transcriptional terminators depends on at least two further factors: (i) there should be no translation through the corresponding segment; (ii) there should be a suitable downstream sequence (T-rich, or resembling the CAATCAA consensus). Unfortunately, even with these criteria, it is difficult to predict with confidence that the putative terminator will actually function. Moreover, all available computer programmes, to my knowledge, depend upon initial identification of inverted-repeat sequences; yet the trpt' terminator does not possess such a feature. (Indeed, none of the features which make this a terminator are at all clear.) Terminators similar to trpt' will therefore, be missed by the computer. Finally, prediction of the stability of possible stem-loops is achieved using rules derived from those of Tinoco et al (1973). Recent evidence confirms a long standing suspicion that these rules leave something to be desired, essentially because they are based on studies of an insufficient range of different synthetic oligonucleotides. One obvious defect of "Tinoco's rules" is that they ascribe essentially no stabilising influence to the G:U base pair. The existence of G:U base pairs, first suggested by Crick (1966) to explain the degeneracy of the genetic code, has now become an accepted feature of RNA secondary structure. The confirmed existence of the G4-U69 base pair in $tRNA^{Phe}$ (Quigley et al, 1975; Ladner et al, 1975) supported the hypothesis of Lomant and Fresco (1975) that G:U and G:T base pairs could occur in helical structures containing predominantly Watson-Crick base pairs. However, unequivocal proof

of G:U base pair formation required the discovery of hydrogen-bonded imino-ring proton resonances in the ^{1}HNMR spectra of tRNA (Johnston and Redfield, 1978, 1981; Robillard et al, 1976). Further support came from NMR- and ethidium bromide binding-studies of poly [d(GpTp)], which found that the polymer formed a stable double-helix (Early et al, 1978). Finally, the existence of G:U base pairs in the double-helical regions of tRNA was firmly established by NMR- and nuclear Overhauser effect-studies (Reid et al, 1979; Hurd and Reid, 1979; Johnston and Redfield, 1978, 1981).

What degree of stabilisation does a G:U base pair provide to an RNA double-helix? This appears to depend on adjacent sequences. Uhlenbeck et al found that terminal G:U base pairs were as stable as terminal A:U base pairs. Clarke (1977) suggested that internal G:U base pairs could be accommodated in an RNA double-helix without causing gross helical distortion. He suggested, however, that such a base pair is a centre of potential weakness, as it requires a certain accommodation of the helix backbone if two H-bonds are to be made. Two lines of evidence provide a more direct insight into the influence of internal G:U base pairs on RNA secondary structure. An NMR-study of a short RNA duplex (CpApUpUpGp:CpApGpUpGp) indicated that such a base pair did not stabilise the duplex, and was in fact a distinct centre of instability (Romaniuk et al, 1979). An extension of this work studied the RNA duplex ApGpGpCpUp:ApGpUpCpUp, where the G:U base pair is flanked not by A:U but by G:C base pairs (Alkema et al, 1982). G:C base pairs are considerably stronger than A:U base pairs, and were expected to allow less conformational perturbation of the helix backbone. Here, the G:U base pair was found to contribute an increase of duplex stability, comparable with that found for an internal A:U base pair. Hence, it seems clear

that G:U base pairs can stabilise RNA secondary structure in some sequences.

Recent studies confirm that the influence of all non-bonded base oppositions is, as suggested by Tinoco et al (1973), disruptive. However, the severity of the effect depends on the bases involved, as does the possibility that more exotic secondary conformations may be formed. Some authors suggest that A.A, C.A and C.C base oppositions can be accommodated within a standard RNA duplex without looping out (Alkema et al, 1982; Lomant and Fresco, 1975). It is generally accepted, however, that all other mismatches take up an extra-helical conformation (for review see Lomant and Fresco, 1975). Tinoco et al considered that all unpaired base oppositions would affect the stability of an RNA structure equally, and so assigned to them all a penalty of +2 Kcal when calculating the free energy of potential structures. Clearly, this may not be justified if A.A, C.A and C.C base oppositions are indeed able to stack in an RNA double-helix without appreciable backbone distortion. The true penalties in these cases may be negligible, even if they contribute no extra stability.

The rules of Tinoco et al have undoubtedly been of value in RNA secondary structure studies. However, it is now evident that predictions based on them may be inaccurate and in some cases midleading. Further research is required before accurate estimates of the stability of potential RNA secondary structures can be made. I have, in practice, found the current computer approach to prediction of terminators unhelpful and shall not discuss it further in the present context. Rather, I will identify sequences (without calculating free energies) which are similar to other known terminators, allowing both G:U pairing and single, non-bonded base oppositions.

Viewed in this way, pIG125 (Fig. 4.7) possesses three reasonably stable "hairpins", one of which is the known tec1 stem-loop (309-316 and 321-328). Termination here is known to be rho-independent. Two-thirds of the *in vitro* transcripts end in a C (position 335) and the remaining one-third end in the following G (position 336) (Dunn and Studier, 1980). Termination occurs in a U-rich region (7/11), although there are no more than 3 consecutive U-residues. As might be expected, this sequence shows no homology to the CAATCAA consensus (i.e. the rho-dependent terminators, λt_{R1}, $tRNA^{Tyr}$ and λt_{RO}). Upstream of tec1, another possible hairpin with a 5 bp stem and a 6 nucleotide loop lies within the cloned fragment (117-121 and 128-132). However, the sequence following this stem has only 5/11 T-residues, and shows no homology with other known terminators suggesting that it may not have termination potential. The third possible hairpin overlaps the right-hand BamHI site and has a stem of 6 bp with a loop of 2 nucleotides (500-505 and 508 and 513). Following this, the sequence CAATAA bears some resemblance to the CAATCAA consensus. However, the smaller tec1 fragment in pIG139 possesses only the tec1 sequence, yet N100 strains harbouring pIG125 and pIG139 exhibit similar efficiencies of termination. It, therefore, seems likely that the termination observed in these strains is due solely to the tec1 sequence (although we cannot formally exclude other sequences present in the pIG139 insert). The other two hairpins present in the pIG125 insert are probably inactive.

Consideration of the predicted translational properties of the cloned sequences casts further light on the situation. In pIG125, ribosomes should enter the mRNA corresponding to the cloned fragment out of phase and terminate translation 218 nucleotides upstream of the normal T7 gene 1.3 stop codon. In pIG139, however, translation

Fig. 4.7

The predicted nucleotide sequence around the insert in pIG125 is shown. The symmetries noted in the text are underlined with arrows. Translation entering the insert is marked by "⌐——¬ " and is termed phase 1. Translation stop codons are underlined and their phase is recorded in brackets.

Fig. 4.7 T7 tec1 in pHR9 ⟶

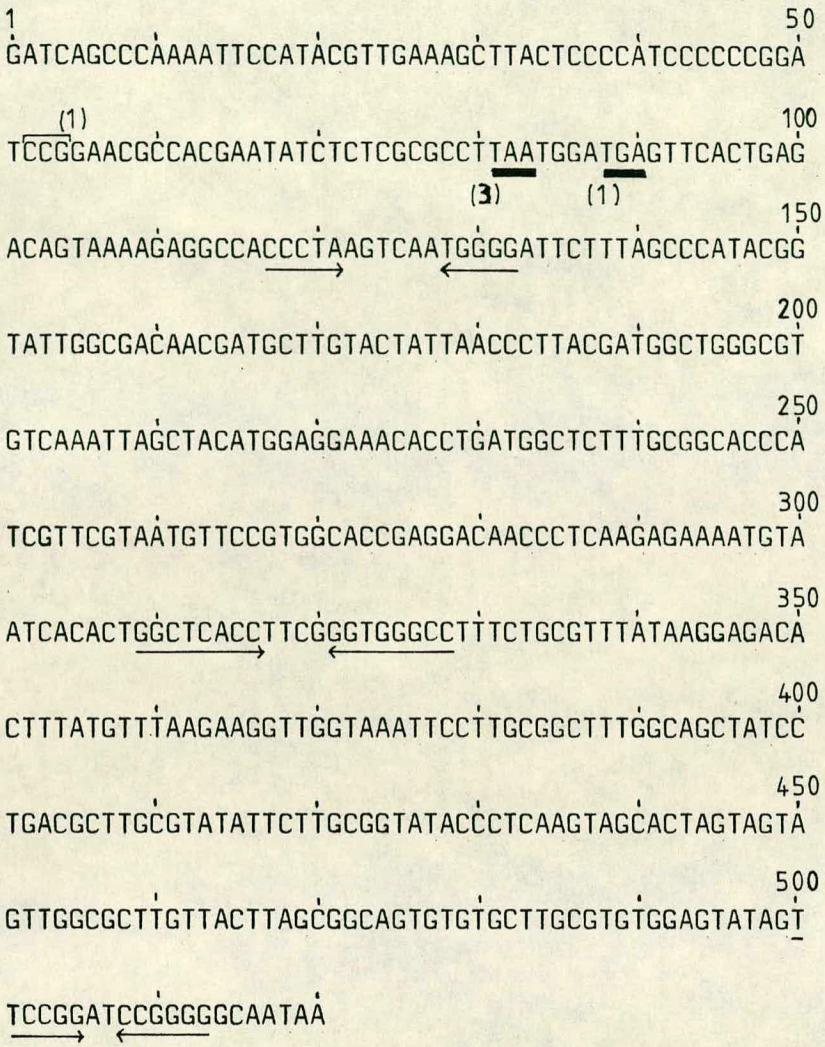


Fig. 4.8

The predicted nucleotide sequence around the insert in pIG126 is shown. Details are marked as in Fig. 4.7.

Fig. 4.8 T7 tec1 in pHR9 ←

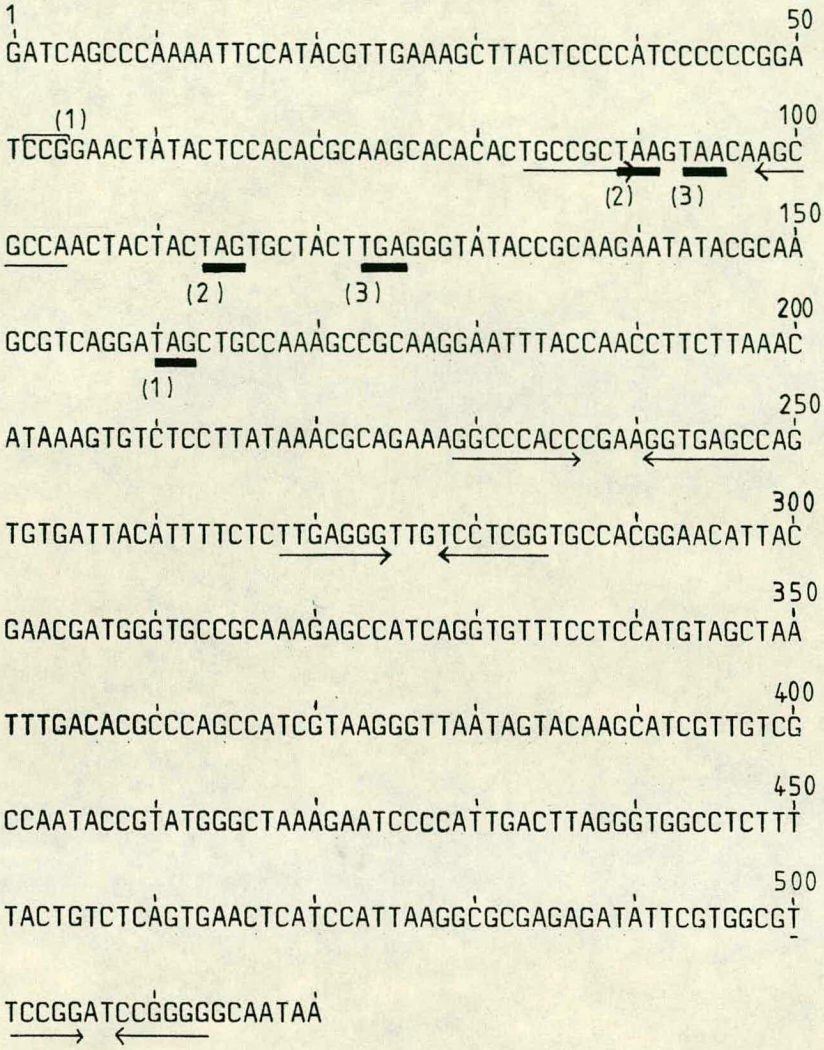


Fig. 4.9

The predicted nucleotide sequences around the inserts in pIG139(a) and pIG138(b) are shown. Details are recorded as in Fig. 4.7.

Fig. 4.9. T7 tec1 (142 bp) fragment in pHR9 ⟶

(a)

1 50
GATCAGCCCAAAATTCCATACTTTGAAAGCTTACTCCCCATCCCCCCGGA

100
TCCAGATCCGGGGCACCCATCGTTCGTAATGTTCCGTGGCACCGAGGACA

(2)

150
ACCCTCAAGAGAAAATGTAATCACACTGGCTCACCTTCGGGTGGGCCTTT

(1)

200
CTGCGTTTATAAGGAGACACTTTATGTTTAAGAAGGTTGGTAAATTCCTT

250
GCGCCGGATCTGGATCCGGGGGCAATAAGGGCTGCACGCGCACTTTTATC

N.B. Stops at same translation stop as T7 in vivo.

(b) T7 tec1 (142 bp) fragment in pHR9 ⟵

1 50
GATCAGCCCAAAATTCCATACGTTGAAAGCTTACTCCCCATCCCCCCGGA

(1) 100
TCCAGATCCGGCGCAAGGAATTTACCAACCTTCTTAAACATAAAGTGTCT

(1) (1) 150
CCTTATAAACGCAGAAAGGCCCACCCGAAGGTGAGCCAGTGTGATTACAT

(1) 200
TTTCTCTTGAGGGTTGTCCTCGGTGCCACGGAACATTACGAACGATGGGT

250
GCCCCGGATCTGGATCCAGATCCGGCGCAAGGAATTTACCAACCTTCTTA

300
AACATAAAGTGTCTCCTTATAAACGCAGAAAGGCCCACCCGAAGGTGAGC

350
CAGTGTGATTACATTTTCTCTTGAGGGTTGTCCTCGGTGCCACGGAACAT

400
TACGAACGATGGGTGCCCCGGATCTGGATCCGGGGGCAATAAGGGCTGCA

CGCGCACTTTTATC

N.B. Three stops before inverted tec1 stem.

should enter the insert in phase, and therefore terminate at the normal place (1.3 gene stop). My results indicate that this difference has little or no effect on the efficiency of termination, suggesting that the final position of upstream ribosomes on the transcript does not influence termination at tec1 significantly. Note that translation is unlikely to reinitiate upstream of tec1 on the pIG125 mRNA. However, ribosomes may well initiate at the putative gene 1.4 start (nucleotide 355: Boothroyd and Hayward, 1979) in the mRNA derived from both plasmids. This may explain why the potential terminator downstream of tec1 in pIG125 is apparently not functional.

Four potential hairpins arise from the pIG126 insertion, only one of which shows downstream sequence homology with known terminators. The first has a 7 bp stem (including a C.C mismatch) and an 8 nucleotide loop (83-89 and 98-104). The second is the inverted tec1 sequence with its internal C.A mismatch (229-236 and 241-248). The third has a 7 bp stem and a 3 nucleotide loop (268-272 and 278-284). The fourth is identical in sequence to the third in pIG125 (500-505 and 508-513) and is followed by CAATAA which resembles the CAATCAA consensus. The first structure is in a translated region, and will therefore probably not function as a terminator. Translation will stop 56 nucleotides upstream of the inverted tec1 sequence. Assuming that translation does not restart the remaining three structures might all contribute to transcriptional termination. The 98% termination observed in N100 (pIG126) must be due to the presence of more than one region, as in N100 (pIG138) we see only an estimated 60% termination per fragment. The insert in this case includes only the second and third possible hairpins. Thus it seems possible that the fourth hairpin of pIG126 is responsible for efficient transcriptional

Fig. 4.10

The predicted nucleotide sequence around the insert in pIG134 is shown. Details are marked as in Fig. 4.7.

Fig. 4.10 T7 tec2 in pHR9 ⟶

termination (just beyond the end of the inserted sequence). Although an identical sequence seemed not to function in pIG125, we have seen that this could reflect translational interference in that case. Finally, the results for pIG138 indicate that weak termination of transcription (totalling 60%) may be generated by one or both of the other potential hairpin structures (the second and third noted above for pIG125) despite their lack of convincing downstream "terminator-tails".

The most surprising observation in this chapter was the failure of the insert in pIG134 to cause any reduction of galactokinase expression. There is evidence that transcription by E.coli Rpol terminates near position 28% on the T7 genome, as well as at tec1 (Minkley and Probnow, 1973; Peters and Hayward, 1974). The BamHI fragment contained in pIG134 is thought to originate from the region which should contain this termination site (Dunn and Studier, 1981, 1982). However, some doubt exists as to whether the fragment has been correctly identified. According to Studier and colleagues, it should also encode an RNAse III target located just ahead of gene 3.8. Nevertheless, they have been unable to demonstrate cleavage by this enzyme of RNA transcribed from the plasmid containing the fragment (Dunn and Studier, 1982). However, my restriction data are consistent with predictions from the sequence they have determined for this fragment. I must, therefore, assume that it has been correctly identified. The fragment (Fig. 4.10) contains twelve regions of significant inverted-symmetry, the strongest of which is that suggested by Hayward (1981) as a potential tec2 site (567-578 and 584-595). It lies in an intercistronic region (between genes 3.5 and 3.8), shows stem-loop/3'-sequence homology to the tec1 sequence (including a similar distribution of U-residues: 7/12) and has a

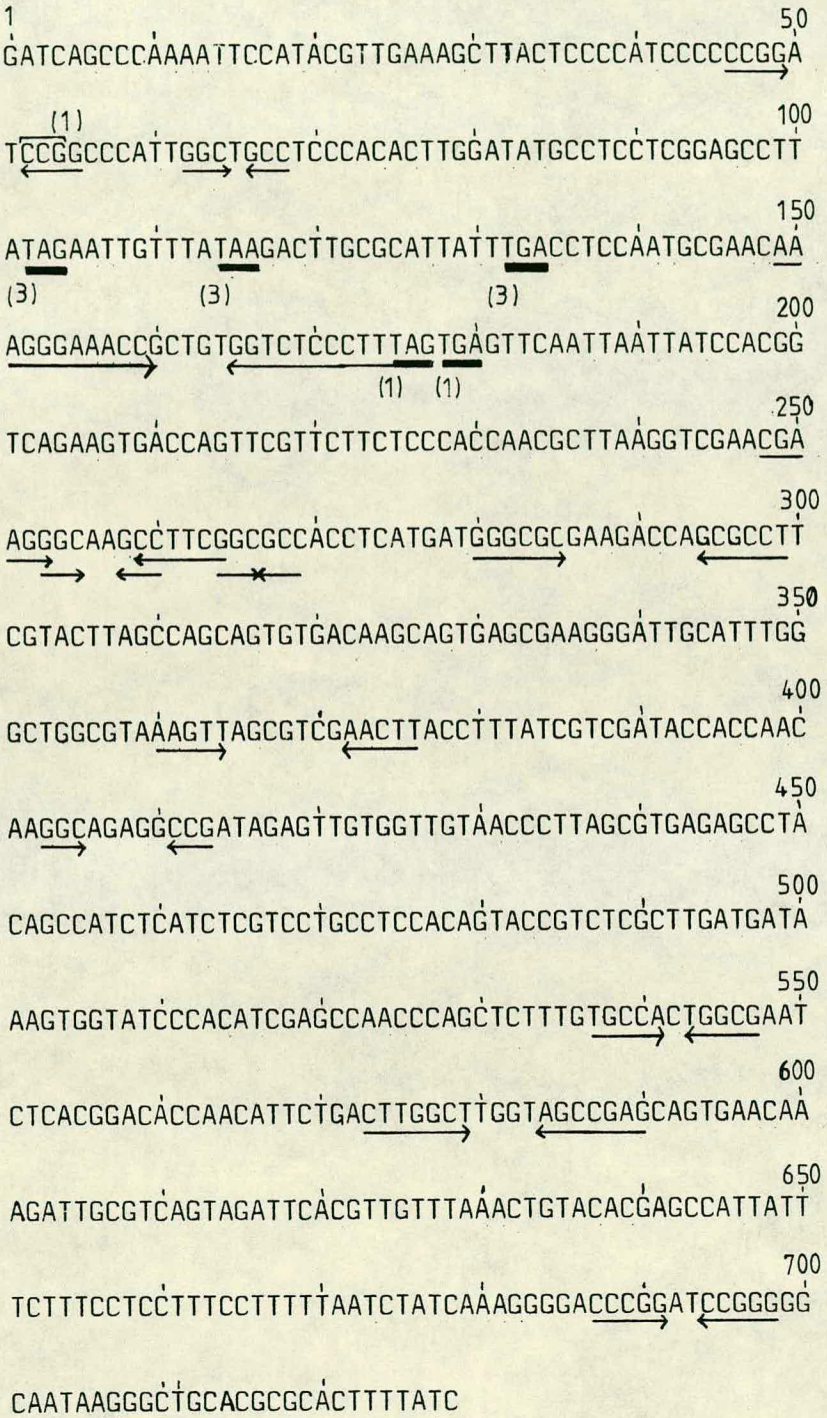
sequence GCA appropriately positioned (in comparison to tec1) to allow termination. Termination here would be expected to be mainly at a C preceding an A, thus explaining the dinucleotide sequence result of Peters and Hayward (1974). Other symmetries occur at: 152-157 and 161-166; 197-199 and 203-205; 331-333 and 338-340; 336-340 and 346-350; 445-449 and 458-462; 477-479 and 483-485; 531-535 and 540-544; 679-681 and 686-688; and lastly, 688-691 and 694-697. Despite the abundance of inverted-symmetries in this fusion, termination was not observed. It was expected that the fragment should contain the tec2 termination site and that this at least, would cause some reduction in galactokinase expression. Translation should be unable to interfere with operation of the proposed tec2 terminator (even with an aberrant "restart" translation upstream) because there are stop codons in all 3 reading frames immediately before tec2 (540-542; 550-552; 554-6; and 558-560). Perhaps 567-595 is not the genuine tec2 stem-loop. If so, it seems unlikely that this terminator exists on the fragment cloned; the nucleotides up to number 94 are part of T7 gene 3; 97 to 549 constitute gene 3.5; and gene 3.8 starts at 616 and continues beyond the end of the cloned fragment (Dunn and Studier, 1981). Thus 567-595 has the only plausible stem-loop potential (in the cloned DNA) which lies in an untranslated region of the original genome. Perhaps tec2 requires additional sequences not present in this insert, and/or a particular Rpol-cofactor interaction, for efficient termination. Finally, it is possible that the tec2 sequence is not located on this fragment at all, but in a region adjacent to it in the T7 genome.

A further surprising result is that this fragment in the inverse orientation (pIG135) reduces galactokinase expression by 99%. The

Fig. 4.11

The predicted nucleotide sequence around the insert in pIG135 is shown. Details are recorded as in Fig. 4.7.

Fig. 4.11 T7 tec2 in pHR9 ←



fusion (Fig. 4.11) contains twelve inverted-symmetries, and one A/T rich region (similar to trpt') with 28 A:T in 40 bp (640-680). The symmetries are in the following regions: 46-49 and 52-55; 62-64 and 66-68; 149-159 and 165-175; 248-253 and 259-264; 253-255 and 258-260; 264-266 and 267-269; 280-285 and 294-299; 360-364 and 372-376; 403-405 and 411-413; 537-541 and 543-547; 573-579 and 584-590; and lastly, 687-691 and 694-698. Translation from the galE region of mRNA would be expected to stop at position 177, where two tandem stop codons occur. This would probably prevent the formation of the first three potential mRNA stem-loops. There is no obvious restart sequence for translation downstream, so the remaining stem-loops may well form. Termination is presumably arising from one or more of these inverted-repeats, and/or the A/T rich region noted earlier.

Concluding this consideration of nucleotide sequence, it is evident that I have only been able to speculate on the sequences involved in any observed reduction of galactokinase expression (presumably through transcriptional termination). Those terminators suggested must be viewed as tentative. Note, however, that the majority of the inverted-repeats discussed lack 3'-sequence homology with other known terminators. Furthermore, many would be present on the untranscribed strand of T7, and would therefore not be expected to function during phage infection. Should future in vitro experiments succeed in defining the 3'-ends of transcripts derived from these plasmids, they may provide useful information regarding "novel" termination signals.

The observed 95% reduction in galactokinase levels arising from the insert in pIG116 suggests that the reduced lacZ expression characteristic of pIG103 does indeed reflect transcriptional termination.

Why then should we see a greater termination efficiency in N100 (pIG116) than in MC1000 (pIG103) (98% versus 78%)? This may reflect translational interference with the termination event in MC1000 (pIG103). On the other hand, insertion of a fragment in pHR1 could relieve polarity inherent in the vector, partially cancelling the effect of the newly introduced terminator. These possibilities serve to stress the inadequacies of the pMC81 system when compared to the pKO system. Only the latter will be considered from here on. It should, however, be noted that the difference in promoters between pIG116 and pIG103 could be relevant for terminator efficiency (c.f. Howe et al, 1982).

Inversion of the tec3 fragment results in normal levels of galactokinase, suggesting that termination does not occur in this orientation. In light of other results, this is unusual. Most fragments show some degree of termination in both orientations. The special properties of the tec3 fragment may become understandable when we consider the nucleotide sequence of the insert and its flanking regions (Chapter 5).

Footnote: A "nusA box" has recently been identified as a consensus sequence with which pnusA is presumed to interact (D. Friedman, pers. comm.). The sequence CGCTCTTA occurs just upstream of phage lambda's N protein targets (λnutL and λnutR) and (with slight variations) at Ø21nutR and P22nutL. Closely related sequences are observed near trpt (CGCAGTTAA, also GGCATTTTAA near the point of termination); trpt' (TGCGCTTAA) and the rrnB "attenuator" (TGCTCTTTAA). Examination of the T7 DNA sequence (Dunn and Studier, 1981) and the sequences presented here reveals the following: a possible "nusA box" at 812-821 nucleotides in the T7 genome (CGCTCTTTAA) just downstream

of E.coli Rpol promoters A_1, A_2 and A_3; two possible "nusA boxes" near the tec1 sequence, one just upstream of it, at nucleotides 74-83 in Fig. 4.7 (CGCGCCTTAA) and one overlapping the point of termination, nucleotides 332-340 in Fig. 4.7 (TGCGTTTA); and one possible "nusA box" within the A/T rich region noted in pIG135 (Fig. 4.11) (GCCATTA). We can, therefore, reconsider the questions raised at the start of this discussion concerning termination at tec1. It seems plausible that pnusA associates with E.coli Rpol, near the A-promoters of T7, and accompanies it during its transcription through the extensive "early" region (see Greenblatt and Li, 1981, for the current model of nusA protein function). Should this be the case, the sequences near tec1 may favour termination by acting as auxiliary "loading/ reloading" sites, or conceivably as "recognition sites". Such sequences might be required if the efficiency of termination is controlled passively by detachment of pnusA from some transcribing complexes, or actively by interaction of pnusA with the "nusA box" such as to cause a conformational change in the attached Rpol, stimulating termination. It should be noted that there is a lack of nonsense polarity in the T7 early operon. In view of this, an antitermination action (possibly via pnusA) seems highly plausible and therefore an "unloading" or a "conversion to termination factor" site might well be required near tec1. However, the first potential "nusA box" is not present in pIG139, yet the inserts in pIG125 and pIG139 produce a similar termination efficiency. It remains plausible that the potential "nusA box" overlapping the termination site is required for termination by one of the means discussed above. These observations also suggest

another possible reason for the failure of the putative tec2 terminator to function in N100 (pIG134). Perhaps pnusA interactions well upstream of this terminator at targets not included in pIG134 are necessary for termination at tec2? It should be noted that a potential "nusA box" (AGCGAATTA) occurs early in Pgal initiated mRNA (Merril et al, 1981).

CHAPTER 5 - DETERMINATION OF THE LOCATION OF tec3 IN THE T7 GENOME AND OF ITS DNA SEQUENCE

5.1 Introduction

The fact that the tec3 fragment appears to cause termination of E.coli Rpol-mediated transcription in pIG103 and pIG116 raises two questions:

i) From what part of the T7 genome does this fragment originate?

ii) What is its nucleotide sequence?

The former question is important as the tec3 fragment is not predicted by the known nucleotide sequence of the first 12,000 bp of the T7 genome (Dunn and Studier, 1981) which includes all regions of the genome normally transcribed by E.coli Rpol. Mapping of the tec3 fragment might therefore suggest other interactions of E.coli Rpol with the phage DNA, apart from its transcriptional activities in the 'early' region. Furthermore, the nucleotide sequence of the tec3 fragment may suggest the region in which termination is occurring and should allow comparison of this sequence with known terminator sequences. In this chapter, I shall describe the determination of: (i) the nucleotide sequence of the tec3 fragment; (ii) its map position; and (iii) its orientation.

5.2 Determination of the nucleotide sequence of the tec3 fragment

Of the nucleotide sequencing methods that have been developed so far, one of the most simple, rapid and accurate is the "M13 dideoxy" method developed by Sanger. This technique has been discussed at length in the literature, and will simply be outlined here. It makes use of the ability of E.coli DNA polymerase to copy a single-stranded DNA template faithfully, using a short DNA fragment as primer; and it exploits the E.coli,male-specific, single-stranded,filamentous

DNA phage M13. Foreign DNA can be cloned into the double-stranded Replicative Form (RF) of one of the genetically engineered vector phages available. The insertion inactivates a segment of the β-galactosidase gene contained in each vector. This segment allows the development of a Lac^+ phenotype in lacZ hosts containing an F-prime carrying a second, complementary segment of the β-galactosidase gene. Strains containing vector molecules produce blue plaques on XG/IPTG plates, whereas those containing recombinant phage give rise to colourless plaques.

After infection, a single-stranded form of the phage (always the positive strand) is extruded without cell lysis. This is easily purified and, using a single, small DNA primer (suitable for all inserts) can be used to make radioactive, complementary copies of the inserted DNA segment. The large fragment of DNA polymerase I ("Klenow enzyme"), which lacks the 5'→3' exonuclease activity of the intact molecule, is used to make these copies. DNA synthesis is carried out in four separate reactions in the presence of all four deoxyribonucleoside triphosphates (one or more of which is ^{32}P-labelled) but with only one of the four dideoxyribonucleoside triphosphates in each reaction. Once the analogue is incorporated, the 3'-end is no longer a substrate for chain elongation and the growing chain is terminated. Hence, each reaction provides a population of partially synthesised, radioactive DNA molecules each having a common 5'-sequence and 3'-terminal base but each varying in length. Parallel size fractionation of the four reaction products on a denaturing polyacrylamide gel and autoradiography allows visualisation of the various DNA chains. The DNA sequence can be read as a ladder representing the shortest chains (5'-end) at the base of the gel to the longest chains (3'-end) at the top of the gel.

The resolution of the gel determines the length of the DNA sequence that can be discerned. The technique is summarised in Fig. 5.1.

Accordingly, pIG116 DNA was digested with HindIII and ligated to HindIII digested M13mp2/HindIII DNA (a HindIII or EcoRI vector). Following transformation into JM101, potential recombinants were identified by their colourless plaque phenotype. These were purified, picked in an ordered array to a lawn of JM101 and probed by hybridisation to [^{32}P]-T7 DNA. A possible source of confusion is that M13 preparations generate deletions upon self-ligation, giving rise to colourless phenotypes. Screening by plaque hybridisation ensures that only true recombinants will be selected for sequencing. Single-stranded template DNA purified from 36 recombinant phage was used in 'T-screen' reactions to search for clones containing the insert in both orientations. This entailed using each template in a single base-specific reaction (in this case with ddTTP). Size fractionation of the products on a polyacrylamide gel, followed by autoradiography, allows visualisation of a pattern characteristic of each template, representing the position of every 'T' in the sequence. Two patterns should be found,representing phage containing the tec3 insert in the two possible orientations. However, of the 36 templates screened, all gave the same characteristic pattern; all must have their inserts in the same orientation. Further attempts to find the opposite orientation were not made as I reasoned that the failure to find this among 36 isolates must imply that there is a strong selection against it. Although the reason for this remains obscure, it is a fairly common problem in M13-cloning. It was therefore only possible to sequence the insert using one strand as template, without confirmation from the opposing strand. The determined

Fig. 5.1. Flow diagram summarising the M13 dideoxy sequencing technique.

Fragments can be cloned at the EcoRI or HindIII sites in M13 mp2/HindIII; the sequence around these sites is shown. Following single-stranded template preparation from recombinant phage, a short DNA primer is annealed to the recombinant so as to allow DNA synthesis through the cloning sites. The 3'-termini of the 26 bp 'universal' primer is shown. The primed template is used in four DNA synthesis reactions containing the four dNTPs (one of which is alpha-labelled with ^{32}P), Klenow polymerase, and one of the four ddNTPs. The partially synthesised, labelled DNA chains (each having a common 5'-terminus) are separated on a denaturing polyacrylamide gel. The smallest DNA chains migrate to the bottom of the gel whilst the longer chains migrate through the gel in proportion to their size. The sequence is read as a ladder from the base to the top of the gel.

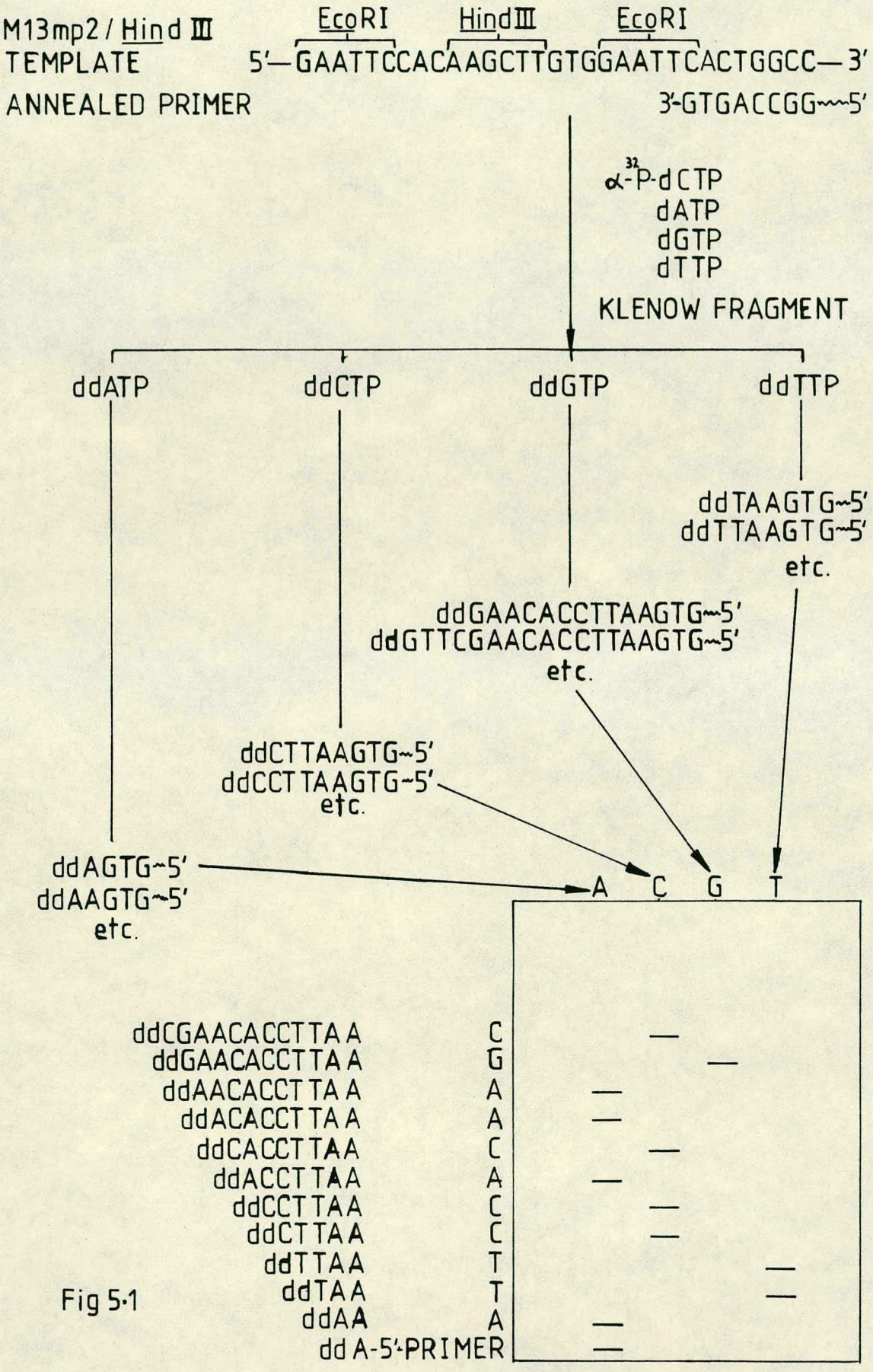

M13mp2 / Hind III
TEMPLATE
EcoRI
HindIII
EcoRI
5'—GAATTCCACAAGCTTGTGGAATTCACTGGCC—3'
ANNEALED PRIMER
3'-GTGACCGG~~5'
α-32P-dCTP
dATP
dGTP
dTTP
KLENOW FRAGMENT
ddATP
ddCTP
ddGTP
ddTTP
ddTAAGTG~5'
ddTTAAGTG~5'
etc.
ddGAACACCTTAAGTG~5'
ddGTTCGAACACCTTAAGTG~5'
etc.
ddCTTAAGTG~5'
ddCCTTAAGTG~5'
etc.
ddAGTG~5'
ddAAGTG~5'
etc.
A C G T
ddCGAACACCTTAA C
ddGAACACCTTAA G
ddAACACCTTAA A
ddACACCTTAA A
ddCACCTTAA C
ddACCTTAA A
ddCCTTAA C
ddCTTAA C
ddTTAA T
ddTAA T
ddAA A
dd A-5'-PRIMER

Fig 5.1

Fig. 5.2

A typical sequencing gel. The position of the tentative HaeIII site is shown. The sequence of the T7 AluI fragment in pIG115-117 reads as follows:

1 60

CTCAGGCCAAGGCTCTGGACGATTCCATGAAGTCAATGAACAAACTTGACGTAATCGACA

T

120

AGCAATTCCAGAAGCGAATCAACGGTGAGTGGGTCTCAACGGATTTTAAGGATATGCCAG

T

TCAACGAGAACATTGGTGAGTTCAAGCATAGCGATATGGTTAACTACGCCAATAAGAAG

Note N/M are possible bases where N is the likelier. In both cases, the sequence of Dunn and Studier (1982) verifies the likelier base. The only difference between my sequence and that of Dunn and Studier (as mentioned in Chapter 5, discussion) is that I assign a T-residue at position 133 whereas they assign a C-residue at this position.

Tracks 1-5 define the position of the HaeIII site within the HindIII insert.

Track	DNA	Restriction Digest
1	pIG115	HaeIII
2	pIG115	HaeIII/HpaI
3	pBR322	HaeIII
4	pIG115	HpaI/HindIII
5	pHR1800	HaeIII

The marker digests are pBR322 and pHR1800, fragment sizes being recorded in base pairs. The sizes of novel fragments in pIG115 digests are shown. This data is consistent with the HpaI site being centred 155 bp from the HaeIII site.

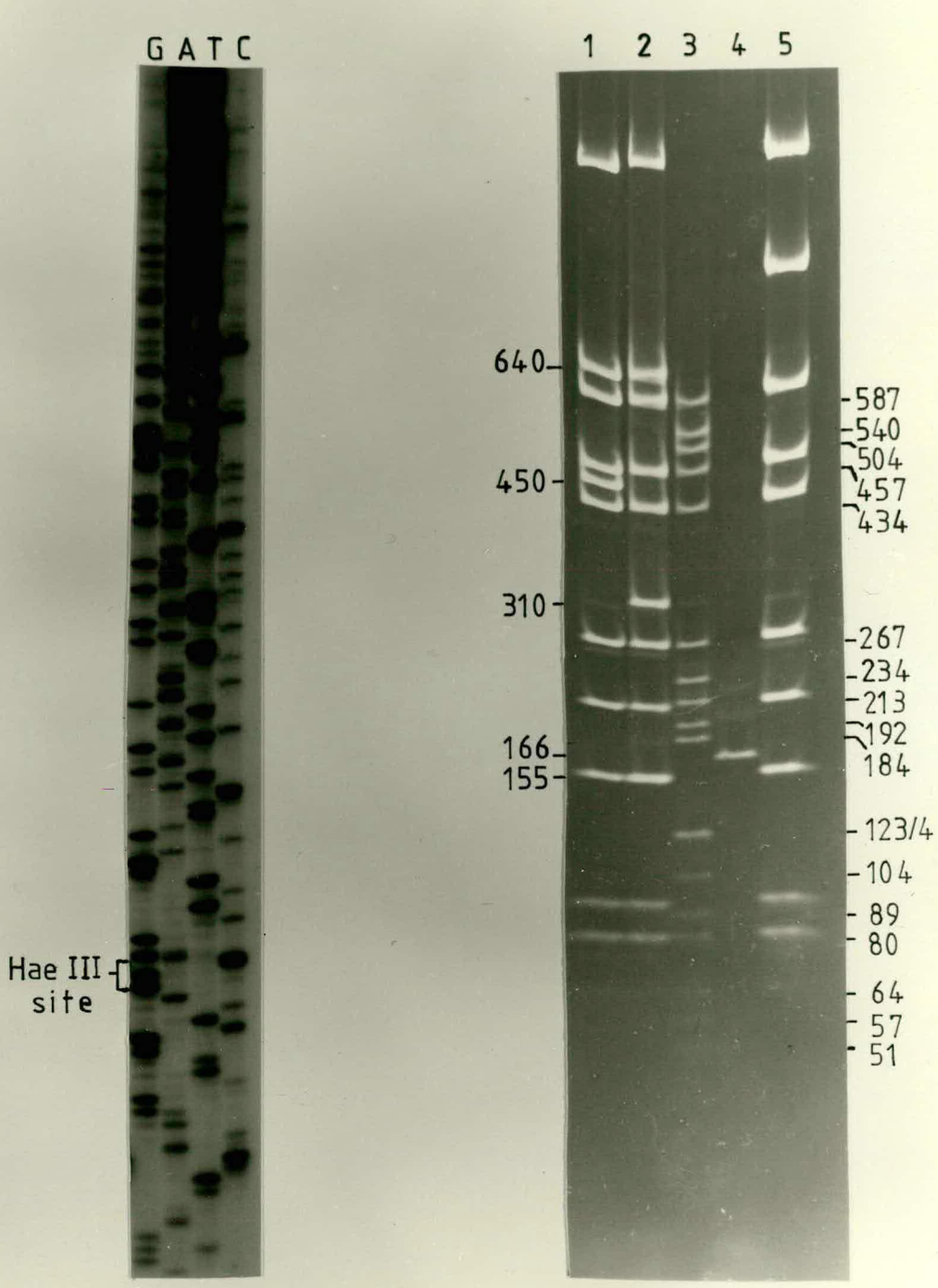

G A T C
Hae III site
1 2 3 4 5
640
450
310
166
155
587
540
504
457
434
267
234
213
192
184
123/4
104
89
80
64
57
51

Fig 5·2

sequence predicts that the HindIII fragment is 189 bp and the AluI fragment 179 bp, with an internal HaeIII site centred 11 bp from the middle of the HindIII site at one end of the fragment. It also predicts an internal HpaI site, 155 bp from the HaeIII target. The HaeIII site was tentative as G-C rich sequences are often difficult to read in M13 sequencing. Its position was verified by HpaI/HaeIII digestion of pIG115 (Fig. 5.2), and proved useful in determining the orientation of the fragment.

5.3 Determination of the location of tec3 in the T7 genome

This was achieved by hybridisation of a [^{32}P]-pIG116 DNA probe to various restriction digests of T7 DNA. The enzymes HpaI, KpnI, AvaI, HindII and HaeII produce a wide size range of T7 fragments (Rosenberg et al, 1979; Boothroyd, 1979). Digests of T7 DNA generated by these enzymes were fractionated by 1% agarose gel electrophoresis and transferred to a nitrocellulose filter by the procedure of Southern (1975). Hybridisation of a [^{32}P]-pIG116 DNA probe to this filter revealed that the plasmid had homology with HpaI "E", KpnI "A", AvaI "E", HindII "D2" and HaeII "D" (Fig. 5.3). The control hybridisation of [^{32}P]-pHR1800 DNA to an equivalent filter showed that the vector molecule did not hybridise to any T7 DNA fragments.

These results strongly suggest that the tec3 fragment contained in pIG116 originates from a region of approximately 675 bp internal to the T7 gene 15 (T7 co-ordinates 72.4% to 74.09%). This was somewhat surprising in that the fragment appears to lie within the coding sequence of a coat core protein, in a region transcribed by T7 Rpol. Furthermore, the occurrence of an internal HpaI site 155 bp from the HaeIII site and the fragment's observed homology with T7 HpaI "E" suggests that this fragment overlaps the HpaI site at the

Fig. 5.3 1% Agarose Gel Electrophoresis

The T7 digests hybridised to [^{32}P]-pIG116 DNA are shown (tracks 1-6). Tracks 7-12 show the resulting autoradiograph revealing fragments of T7 with homology to pIG116. The relevant fragments are recorded.

Track	DNA	Restriction digest
1	pHR1800	EcoRI
2	T7	KpnI
3	T7	HindII
4	T7	AvaI
5	T7	HpaI
6	T7	HaeII
7	pHR1800	
8	T7 KpnI "A"	
9	T7 HindII "D2"	
10	T7 AvaI "E"	
11	T7 HpaI "E"	
12	T7 HaeII "D"	

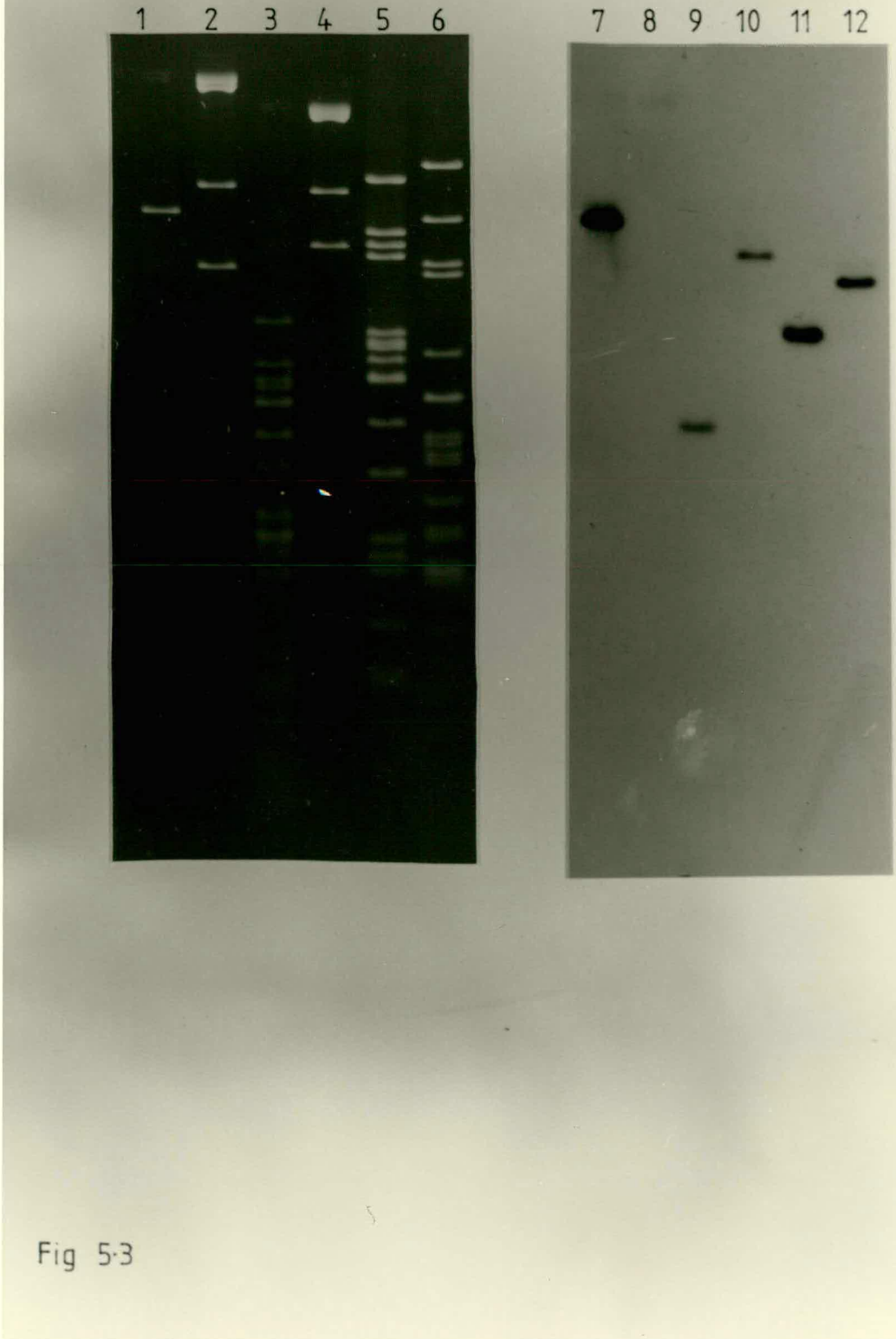


Fig 5·3

right end of HpaI "E", and lies mostly within HpaI "E". If so, the fragment in pIG103 and pIG116 has a transcriptional orientation opposite to that in T7.

5.4 Determination of the orientation of the tec3 fragment in the T7 genome

To test the conclusion that the HaeIII site in the tec3 fragment lies within HpaI "E", I investigated the position of the HaeIII sites around the HpaI site at 74.09% (Fig. 5.4). Restriction data from this region predict that an AvaI site occurs approximately 497 bp downstream of the HpaI target (i.e. in HpaI "P"). This Ava I site is at the right-end of AvaI "E", which I purified from a 1% agarose gel following electrophoresis. The 5′-termini of the fragment were labelled with ^{32}P using polynucleotide kinase, and the fragment was digested with HaeII to separate the labelled ends. The left-hand heterospecific fragment is about 880 bp, and the right-hand one is about 1160 bp. These were separated by electrophoresis on a 5% polyacrylamide gel and the larger labelled fragment was purified by passive elution. This heterospecific fragment, with label at one end, was then partially digested with HaeIII. Following electrophoresis on a 3-10% polyacrylamide gel, and autoradiography, two labelled sub-fragments were observed, one of about 1160 bp and one of about 638 bp. This suggests that there is only one HaeIII site in the 1160 bp heterospecific HaeII-AvaI fragment from the right-hand end of AvaI "E". Furthermore, the HpaI site (at 74.09%) is approximately 497 bp and the HaeIII site approximately 638 bp to the left of the labelled AvaI site. This places the HaeIII site nearest to HpaI - 74.09% about 141 bp to its left, and confirms the suspected orientation of the tec3 fragment

Fig. 5.4 3-10% Acrylamide Gel Electrophoresis

Partial HaeIII digestion of the labelled, right-hand, hetero-specific fragment produced on HaeII digestion of 5'-labelled T7 AvaI "E". The marker track is pBR322. The sizes of fragments are recorded in base pairs.

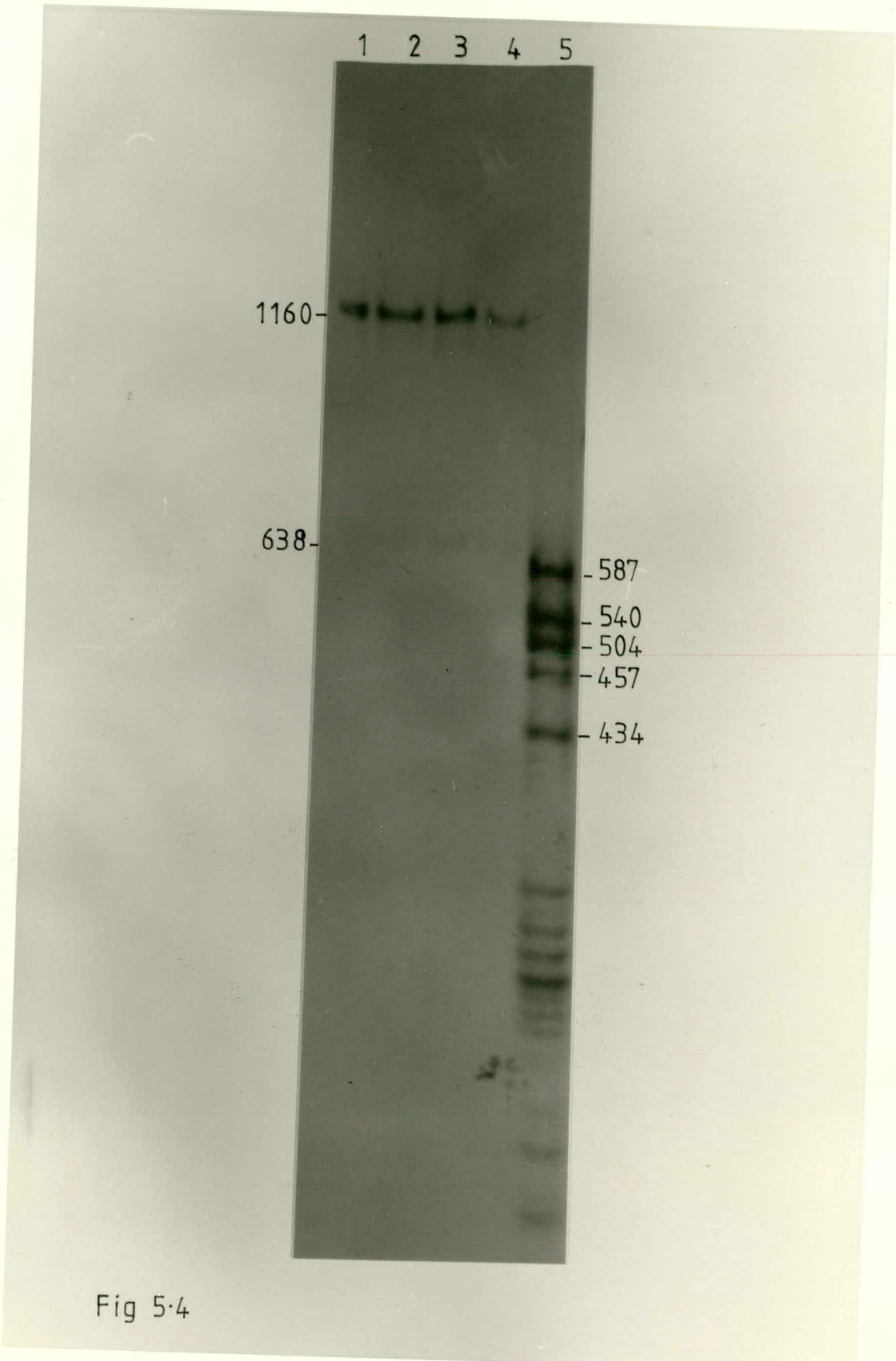


Fig 5·4

in the T7 genome, i.e. opposite to its transcriptional orientation in pIG103 and pIG116.

Finally, it should be noted that the conclusions from sections 5.3 and 5.4, and the sequence determined in section 5.2, are all supported by sequence data received recently from Dr J.Dunn, who has just completed the nucleotide sequence of the whole T7 genome. The few differences between our sequence data are recorded in Fig. 5.2.

5.5 Discussion

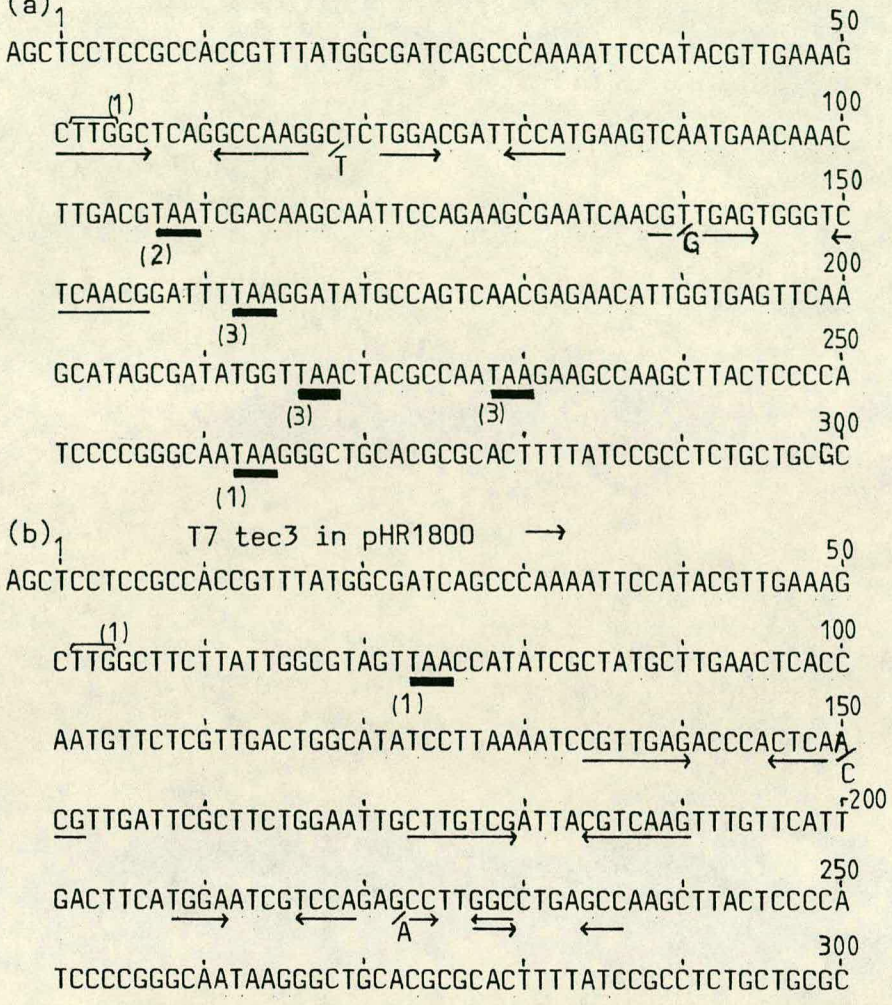
The data presented in this chapter unequivocally map the tec3 fragment within the T7 genome to the region of the HpaI target at 74.09%. The sequence determined for the fragment (Fig. 5.5) can be examined for plausible termination sites, as in Chapter 4. Three inverted-repeats are evident in pIG115: 51-56 and 61-66; 71-74 and 79-82; and 138-144 and 150-156. None of these shows good homology to known transcriptional terminators on their 3'-flanks. Furthermore, ribosomes entering the mRNA corresponding to the insert, from galE', are predicted to continue all the way through the T7 sequence. It is therefore not surprising that galactokinase levels in N100 (pIG115) suggest no termination of transcription by the insert. Even if any of the inverted-repeats were a potential terminator, translation could inactivate it by preventing hairpin formation by the nascent RNA. Studies of the transcription of pIG115 in vitro, in the absence of ribosomes, might clarify this issue.

In pIG116, five inverted-repeats are evident: 134-140 and 146-152; 173-179 and 184-190; 208-211 and 216-219; 222-224 and 227-229; and 227-229 and 234-236. Only the first two symmetries are flanked on the 3'-side by sequences homologous to known terminators; both have 3'-sequences rich in T-residues (6/11 and 7/11 respectively).

Fig. 5.5

The nucleotide sequences around the inserts in pIG115(a) and pIG116(b) are shown. Details are recorded as in Fig. 4.7.

Fig. 5.5 T7 tec3 in pHR1800 ←



Furthermore, the first symmetry has CG positioned similarly to that of tec1 and the second a CA at the same point as in the putative tec2 sequence. Translation of the mRNA corresponding to the insert should terminate at nucleotides 72-75 and would, therefore, probably not interfere with hairpin formation by any of the five symmetries. In particular, restarting of translation before the first two sites seems unlikely. The observed 95% reduction in galactokinase expression ascribable to the insert in pIG116 could be due to transcriptional termination at one or more of these sites. The best are the first two symmetries, in view of their 3'-flanking sequences. Future experiments, designed to define the 3'-terminus of transcripts derived from these plasmids in vitro, should resolve this uncertainty. They may also reveal termination within the pIG115 insert in the absence of translation as discussed previously.

The mapping of this fragment reveals that the observed termination is the result of yet another signal so orientated as to affect leftward transcription of T7 DNA by the host Rpol if this actually ever occurs. Terminators seem quite common in this orientation.

It should be noted that my sequence differs at only one point from that determined by Dunn and Studier (1982). I assign an A-residue at position 102, where they propose a G-residue. This may reflect a genuine mutation, which would place an isoleucine instead of a threonine at the corresponding position in p15. This could well be a "silent" mutation.

The nucleotide sequence also confirms my earlier restriction data. The HaeIII cut is located 11 bp from the centre of the nearest HindIII site, and the HpaI site is centred 155 bp from the HaeIII cut. My estimates of 12 bp and 155 bp are, thus, in good agreement with the sequence data.

CHAPTER 6 - FINAL DISCUSSION

A major problem in interpreting the data presented here has been the lack of information concerning the stop-sites of terminated transcripts. Elucidation of these *in vitro* and *in vivo* will clarify the situation considerably. In view of the many putative terminators observed, S1-treatment of RNA:DNA hybrids would be my method of choice to achieve this. This method recommends itself in situations where there may well be multiple termination sites and locates the regions in which termination occurs.

My estimates of tec1's termination efficiency in the pKO system are in good agreement with that of Millette *et al* for T7 itself (1970). Thus, termination at this site appears to be 76-78% efficient when measured in isolation from the rest of the T7 genome. This observation contradicts the claim of Ponta *et al* (1974) that the T7 *0.7* product is required for efficient termination at this site during phage infection. However, the "inelegant alternative" discussed earlier (Chapter 4) remains possible.

The failure to demonstrate termination activity in N100 (pIG134) was surprising in that other authors have observed termination *in vitro* in this general region. The reasons for this failure are open to speculation. Perhaps other sequences not present on the tec2 fragment are necessary for an efficient termination at this site *in vivo*? Should my suggestion that a cofactor involvement is necessary for termination at tec1 be true, we may need to mimic this situation when observing tec2 in the pKO system. The observed potential "nusA boxes" in T7 (shortly after the promoters A_1, A_2 and A_3, and in the region of tec1) suggest that *pnusA* may influence transcription of the early region. Should this be the case, we might need to imitate the relevant *pnusA*/Rpol interactions typical

of T7 transcription to observe tec2 stimulated termination in the pKO system. Cloning the tec2 fragment downstream of the tec1 insert (in pIG125 or pIG139) may achieve this. Alternatively, translation may inhibit termination at this site as the previous demonstrations of termination in this general region were achieved in its absence *in vitro*. Translation of RNA transcripts from T7 and pIG134 may prevent termination at tec2. It should be noted that the insert contained in pIG134 contains the complete *3.5* gene (nucleotides 97-550, Fig. 4.10). Translation of RNA encoding the *3.5* gene product may inhibit the termination observed *in vitro*. This possibility would be more plausible should tec2 lie not in the region suggested by Hayward, but within the *3.5* gene. A further possibility is that tec2 lies in an adjacent region to this insert. To resolve this, we require to identify the termination site more precisely, perhaps by studying T7 mutants which lack the tec1 sequence (e.g. T7 LG37) *in vitro* and *in vivo*.

The identification of a new transcription terminator, tec3, (even if it is non-physiological) provides new working material to investigate termination. Its nucleotide sequence reveals potential sites for termination, two of which are similar to tec1 in that they have T-rich 3'-sequences. However, identification of the 3'-terminus of the terminated RNA is essential to identify the termination signal. Further study of termination here (and in the other fusions constructed) may provide insight into termination at "novel" terminators as most of the putative termination sites noted have inverted dyad-symmetries but lack 3'-sequence homology with known terminators. Perhaps the most interesting region noted will be the A/T rich region noted in pIG135. This is similar to *trpt'* in that its A/T content is 70% (63% for *trpt'*) and both contain a potential "*nusA* box".

Should termination occur here, it will be only the second known example of a terminator sequence of this type.

In considering which of the putative terminators are likely to function, it may be enlightening to compare them with mutant and synthetic terminator sequences. Let us first consider the structure of the inverted repeat. The strength of the stem appears to be positively correlated with terminator efficiency. For example: in λt_{R1}, mutations exist that increase (cin) or decrease (cnc) the stability of the stem. The efficiency of termination at these mutant sites increases (cin) or decreases (cnc) depending on the mutation involved. Similarly, in trp att, mutations that increase or decrease trp operon expression have been isolated. Their properties support the importance of the stable formation of stem-loop 3:4 for termination. Increased trp operon expression results from a reduced stability of the 3:4 stem-loop. Decreased trp operon expression results from an enhancement of its formation by abolition of translation or of stem-loop 2:3 formation. The synthetic terminators trpC (RI) and trpC (Bam) (Christie and Platt, 1980; Christie et al, 1982) also show a positive correlation between stem-strength and termination efficiency. The former has a stem of 6 bp (5 G:C), a loop of 4 nucleotides, and encodes a "tail" of 4 U-residues. The latter is identical in these respects but has a stem of 7 bp (6 G:C). Termination in vitro (15%) is only observed in the latter sequence (using wild-type Rpol). Hence the strength of the stem appears to affect termination efficiency. Moreover, the authors suggest that a stem of 8 bp is the maximum useful length.

What then of "tail" 3'-sequences? Most rho-independent terminators produce transcripts ending in a run of U-residues. A notable exception is T7 tec1 which produces transcripts ending at a C- or

G-residue in a U-rich region (7/11). Terminator mutants that have a reduced number of U-residues following the stem-loop show a reduced efficiency of termination. For example: the trp att mutant trpa 1419 has only four U-residues in this region (4/11) compared to eight (8/11) in the wild-type sequence. The altered attenuator is only partially functional in vivo and no termination is detectable in vitro (Bertrand et al, 1977; Platt, 1981). Clearly, the presence of U-residues distal to a stem-loop is important for efficient termination. But is a long run of U-residues more important than a general U-richness? Many terminator sequences (having long runs of U-residues) and the particular instability of polyrU.dA (Martin and Tinoco, 1980) suggest the former. However, Hayward (pers. comm.) has suggested that, in view of such sequences as T7 tec1 and rpoBC att, the latter may sometimes be just as effective. Should this be the case, the putative terminator sites with this structure, which I have proposed above, may indeed be functional. Moreover, a tendency towards U-richness following stem-loops would, of necessity, produce long runs of U-residues.

With these thoughts in mind, few of the putative sites are predicted to be actual sites of transcriptional termination. The efficient termination observed with the inverted tec1 and tec2 fragments is, therefore, of potential interest. Identification of the point(s) of termination in these fusions may identify novel terminator tail sequences, or further examples of the unique trpt' class. Clearly, in vitro experiments with these fusions is necessary to resolve these speculations. It is, however, evident that nucleotide sequence, with or without free energy calculations, is insufficient in the present state of knowledge to identify transcriptional terminators.

Should the reduction in galactokinase levels observed in N100 (pIG126), N100 (pIG138), N100 (pIG135) and N100 (pIG116) be due to transcriptional terminators, how can we explain their occurrence in such an orientation in T7 that they would not affect natural transcription? The established transcription of T7 DNA during infection is confined to the rightward direction (using the r-strand as template), with the possible exception of that initiated at the A_o promoter. However, Rpol does bind to the l-strand *in vitro* when Rpol:DNA ratios are high. Furthermore, promoters for *E.coli* Rpol that could allow initiation of leftward transcription have been isolated (Dunn and Studier, 1982). It is, therefore, possible that transcription initiates on the l-strand to some degree during infection. This would result in convergent transcription between Rpol molecules transcribing left to right and those transcribing right to left. Ward and Murray (1979) investigated this situation using λ*trp* transducing phages. Here, transcripts initiated at the *trp* promoter and λP_L converged and were shown to be mutually impaired. Should this situation occur during T7 infection, it would hinder gene expression. The presence of transcription terminators in the genome that prevent appreciable leftward transcription may help to alleviate such a problem. The competition between opposing transcription complexes is a "dynamic process" (as indicated by Ward and Murray). Prompt termination of leftward transcription may help rightward transcription complexes to displace the opposing complexes.

On the other hand, the presence of these sequence may be fortuitous. As seen in Chapter 4, inverted-symmetries are relatively common in a large nucleotide sequence. These may be present by chance, or for a "purpose": for example, they may promote hairpin formation in transcripts so as to regulate the rate of transcription and/or the spacial

relationship between the polymerase and a following ribosome (recall the "pause" site in the trp leader region, Chapter 1); or they may be protein recognition sequences. The further occurrence of "fortuitous" sequences downstream of the stem-loops could "create" functional transcriptional terminators.

The pKO system can be exploited to investigate termination in a number of ways. It is not yet clear whether the position at which ribosomes halt upstream of a transcriptional terminator will influence termination and, if so, whether terminators differ in this respect. This can be investigated by the construction of similar fusions to those described above but with translation terminating at different sites upstream. To achieve this, it may be necessary to construct new vectors derived from the existing pKO vectors. Moreover, the assumed inactivation of transcriptional terminators by ribosomes translating through them has yet to be demonstrated. This prediction can be tested by constructing a pKO fusion containing a terminator upstream of galK such that translation would extend through the inserted fragment. The effect of translational readthrough on rho-dependent and rho-independent terminators could be tested in this way. Terminator-containing fragments suitable for constructing these fusions are available: McKenney et al (1981) possess a λt_{R1} fragment suitable for the rho-dependent situation; and the 142 bp tec1 fragment described above is suitable for the rho-independent situation.

The pKO system can also be used to study mutant terminator sites. Site-specific mutagenesis, using cloned terminator fragments, will enable the reconstruction of defined pKO fusions containing mutant terminators. Comparison of galactokinase expression by these fusions with that of a fusion containing the wild-type terminator will enable

the influence of each mutation on transcription readthrough to be monitored. Moreover, the presence of each mutant terminator on a small cloned fragment will allow the easy characterisation of each mutation by DNA sequencing. However, as noted earlier (Chapter 4), the screening of resulting clones for mutant-terminators using colour-phenotype may prove difficult. Should tec1 be used in this approach, terminator "up" mutations would reduce galactokinase expression and perhaps produce levels sufficient to cause a noticeable colour-change. It may, nevertheless, be necessary to utilise a weaker promoter than Pgal to achieve the range of galactokinase expression suitable for mutant-terminator screening. This would undoubtedly be necessary when screening for tec1 terminator "down" mutations.

The present studies and considerations suggest a number of avenues by which the investigation of termination could continue: 1) as noted above, in vitro mutagenesis should be used to isolate terminator specific mutants in the pKO system; 2) the present fusions should be transferred to a single-copy when the relevant pKO vectors become available; 3) the point(s) of termination in each fusion described above should be determined in vitro and in vivo; 4) the way in which Rpol interacts with sequences around the termination site should be studied (the tec3 fragment has been sent to Dr S.Beychok, Columbia University, New York, who intends to investigate terminators in this way; 5) the possible function of these terminators in the T7 genome should be studied by using these fragments (in single-stranded form) as probes for RNA produced during infection; 6) the possible involvement of pnusA in events at tec1, tec2 and tec3 should be investigated in N100 nusA strains.

BIBLIOGRAPHY

Adhya, S. and S. Garges (1982) Cell 29, p. 287-289

Adhya, S. and M. Gottesman (1978a) Ann.Rev.Genet. 13, p. 319-533

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

Adhya, S., M. Gottesman and B. de Crombrugghe (1974) Proc.Natl.Acad. Sci.USA 71, p. 2534-2538

Adhya, S., M. Gottesman, B. de Crombrugghe and D. Court (1976) in "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 719-730, Cold Spring Harbor Laboratory, New York

Adhya, S., P. Sarkar, D. Valenzuela and U. Maitra (1979) Proc.Natl. Acad.Sci.USA 76, p. 1613-1617

Alkema, D., P.A. Hader, R.A. Bell and T. Neilson (1982) Biochemistry 21, p.2109-2117

Anderson, S., M.J. Gait, L. Mayol and I.G. Young (1980) Nucleic Acids Res. 8, p. 1731-1743

Armstrong, V.W., H. Sternbach and F. Eckstein (1976) Biochemistry 15, p. 2086-2091

Baker, R. and C. Yanofsky (1972) J.Mol.Biol. 69, p. 89-102

Barry, G., C.L. Squires and C. Squires (1979) Proc.Natl.Acad.Sci.USA 76, p. 4922-4926

Barry, G., C. Squires and C.L. Squires (1980) Proc.Natl.Acad.Sci.USA 77, p. 3331-3335

Bektesh, S.L. and J.P. Richardson (1980) Nature 283, p. 102-104

Benton, W.D. and R.W. Davis (1977) Science 196, p. 180-182

Berg, D., K. Barrett and M. Chamberlin (1971) in "Methods in Enzymology", eds. L. Grossman and K. Moldave. vol. XXI, p. 506-519, Academic Press, New York.

Bertrand, K., L.J. Korn, F. Lee and C. Yanofsky (1977) J.Mol.Biol. 117, p. 227-247

Bertrand, K., C. Squires and C. Yanofsky (1976) J.Mol.Biol. 103, p. 319-337

Bertrand, K. and C. Yanofsky (1976) J.Mol.Biol. 103, p. 339-349

Birnboim, H.C. and J. Doly (1979) Nucleic Acids Res. 7, p. 1513-1523

Blumental, R.M., S. Reeh and S. Pederson (1976) Proc.Natl.Acad.Sci. USA 73, p. 2285-2288

Boothroyd, J.C. (1979) PhD Thesis, Edinburgh University

Boothroyd, J.C. and R.S. Hayward (1979) Nucleic Acids Res. 7, p. 1931-1943

Bronson, M.J., C. Squires and C. Yanofsky (1973) Proc.Natl.Acad.Sci. USA 70, p. 2335-2339

Burgess, R.R. (1976) In "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 69-100, Cold Spring Harbor Laboratory, New York.

Burgess, R.R., A.A. Travers, J.J. Dunn and E.K.F. Bautz (1969) Nature 221, p. 43-46

Calva, E. and R.R. Burgess (1980) J.Biol.Chem. 255, p. 11017-11022

Cameron, J.R., P. Phillipsen and R.W. Davis (1977) Nucleic Acids Res. 4, p. 1429-1448

Casadaban, M.J. (1975) Proc.Natl.Acad.Sci.USA 72, p. 809-813

Casadaban, M.J. and S.N. Cohen (1980) J.Mol.Biol. 138, p. 179-207

Chamberlin, M.J. (1976) In "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 17-67, Cold Spring Harbor Laboratory, New York.

Chang, A.C.Y. and S.N. Cohen (1978) J.Bacteriol. 134, p. 1141-1156

Christie, G.E., P.J. Farnham and T. Platt (1982) in press

Christie, G.E. and T. Platt (1980) J.Mol.Biol. 142, p. 519-530

Clarke, B.F.C. (1977) Prog.Nucleic Acid Res.Mol.Biol. 20, p. 1-19

Coppo, A., A. Manzi, J.F. Pulitzer and H. Takahashi (1975) J.Mol. Biol. 96, p. 601-624

Crawford, I.P., B.P. Nichols and C. Yanofsky (1980) J.Mol.Biol. 142, p. 489-502

Crick, F.H.C. (1966) J.Mol.Biol. 19, p. 548-555

Darlix, J.-L. (1973) Eur.J.Biochem. 35, p. 517-526

Das, A., D. Court and S. Adhya (1976) Proc.Natl.Acad.Sci.USA 73, p. 1959-1963

Das, A., C. Merrill and S. Adhya (1978) Proc.Natl.Acad.Sci.USA 75, p. 4828-4832

de Crombrugghe, B., S. Adhya, M. Gottesman and E. Pastan (1971) Nature New Biol. 241, 260-264

de Crombrugghe, B. and I. Pastan (1978) "The Operon", eds. J.H. Miller and W.S. Reznikoff, p. 304-324, Cold Spring Harbor Laboratory, New York.

Delius, H., H. Westphal and N. Axelrod (1973) J.Mol.Biol. 74, p. 677-687

Dickson, R.C., J. Abelson, P. Johnson, W.S. Reznikoff and W.M. Barnes (1977) J.Mol.Biol. 111, p. 65-79

Drahos, D. and W. Szybalski (1981) Gene 16, p. 261-274

Dunn, J.J. and F.W. Studier (1980) Nucleic Acids Res. 8, p. 2119-2132

Dunn, J.J. and F.W. Studier (1981) J.Mol.Biol. 148, p. 303-330

Dunn, J.J. and F.W. Studier (1982) in press.

Early, T.A., J. Olmsted, D.R. Kearns and A.G. Lezius (1978) Nucleic Acids Res. 5, p. 1955-1970

Farnham, P.J. and T. Platt (1980) Cell 20, p. 739-748

Farnham, P.J. and T. Platt (1981) Nucleic Acids Res. 9, p. 563-577

Finger, L.R. and J.P. Richardson (1979) Fed.Proc.Fed.Amer.Soc. Exp.Biol. 38, p. 403

Franklin, N.C. (1974) J.Mol.Biol. 89, p. 33-48

Friedman, D.I. (1971) In "The Bacteriophage Lambda", ed. A.D. Hershey, p. 733-738. Cold Spring Harbor Laboratory, New York.

Friedman, D.I., A.T. Schauer, M.A. Baumann, L.S. Baron and S.L. Adhya (1981) Proc.Natl.Acad.Sci.USA 78, p. 1115-1118

Friedman, D.I., G.S. Wilgus and R.J. Mural (1973) J.Mol.Biol. 81, p. 505-516

Fujiki, H., P. Palm, W. Zillig, R. Calendar and M, Sunshine (1976) Mol.Gen.Genet. 145, p. 19-22

Fukuda, R. and A. Ishihama (1974) J.Mol.Biol. 87, p. 523-540

Fukuda, R., A. Ishihama, T. Saitoh and M. Taketo (1977) Mol.Gen. Genet. 154, p. 135-144

Fuller, R.S. and T. Platt (1978) Nucleic Acids Res. 5, p. 4613-4623

Galas, D. and A. Schmitz (1978) Nucleic Acids Res. 5, p. 3157-3170

Gallupi, G.R. and J.P. Richardson (1980) J.Mol.Biol. 138, p. 513-539

Gardener, J. (1979) Proc.Natl.Acad.Sci.USA 76, p. 1706-1710

Gilbert, W. (1976) In "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 193-203. Cold Spring Harbor Laboratory, New York.

Gottesman, M., S., Adhya, D. Court and A. Das (1978) Proc.Febs. Mtg. 11th

Greenblatt, J. (1981) Cell 24, p. 8-9

Greenblatt, J., P. Malnoe and J. Li (1980) J.Biol.Chem. 255, p. 1465-1470

Greenblatt, J., M. McLimont and S. Hanly (1981) Nature 292, p. 215-220

Greenblatt, J. and J. Li (1981) Cell 24, p. 421-428

Greenblatt, J., J. Li, S. Adhya, D.I. Friedman, L.S. Baron, B. Redfield, H.-F. Kung and H. Weissback (1980) Proc.Natl.Acad. Sci.USA 77, p. 1991-1994

Gros, G., D.A. Fields and E.M.K. Bautz (1976) Mol.Gen.Genet. 147, p. 337-341

Grunstein, M. and D.S. Hogness (1975) Proc.Natl.Acad.Sci.USA 72, p. 3961-3965

Guarente, L.P. and J. Beckwith (1978) Proc.Natl.Acad.Sci.USA 75, p. 294-297

Guarente, L.P., J. Beckwith, A.M. Wu and T. Platt (1979) J.Mol. Biol. 133, p. 189-197

Gunsalas, R.P., G. Zurawski and C. Yanofsky (1979) J.Bact. 140, p. 106-113

Hausman, R. (1976) Curr.Topics in Microbiol. and Immunol. 75, p. 77-110

Hayward, R.S. (1980) Analyt.Biochem. 108, p. 282-284

Hayward, R.S. (1981) pers. comm.

Hayward, R.S. and G.G. Peters (1974) FEBS Proc. 33, p. 361-366

Heil, A. and W. Zillig (1970) FEBS Lett. 11, p. 165-168

Hercules, K., S. Jovanovich and W. Sauerbier (1976) J.Virol. 17, p. 642-658

Hershey, A.D. (ed.) (1971) "The Bacteriophage Lambda", Cold Spring Harbor Laboratory, New York.

Hesselbach, B.A. and D. Nakada (1977) J.Virol. 24, p. 736-745

Hillel, Z. and C.-W. Wu (1978) Biochemistry 17, p. 2954-2961

Hinkle, D.C. and M.J. Chamberlin (1972) J.Mol.Biol. 70, p. 157-185

Hinkle, D.C., J. Ring and M.J. Chamberlin (1972) J.Mol.Biol. 70, p. 197-207

Hiraga, S., K. Ito, K. Hamada and T. Yura (1967) Biochem.Biophys. Res.Commun. 26, p. 522-527

Hong, J.-S., G.R. Smith and B.N. Ames (1971) Proc.Natl.Acad.Sci.USA 68, p. 2258-2262

Hopkins, A.S., N.E. Murray and W.J. Brammar (1976) J.Mol.Biol. p. 549-569

Howard, B.H. and B. de Crombrugghe (1976) J.Biol.Chem. 251, p. 2520-2524

Howe, K.M., A.J. Newman, I. Garner, A. Wallis and R.S. Hayward (1982) submitted to Nucleic Acids Res.

Hu, S.-L., J.S. Salstrom and W. Szybalski (1979) Virol. 98, p. 411-423

Hu, S.-L. and W. Szybalski (1979) Virol. 98, p. 424-432

Hurd, R.E. and B.R. Reid (1979) Biochemistry 18, p. 4017-4024

Imamoto, F. (1968) Nature 220, p. 31-34

Inouye, M., N. Arnheim and R. Sternglanz (1973) J.Biol.Chem. 248, p. 7247-7252

Ishihama, A., S. Murakami, R. Fukuda, A. Matsukage and T. Kameyama (1971) Mol.Gen.Genet. 111, p. 66-76

Ishihama, A., M. Taketo, T. Saitoh and R. Fujuda (1976) In "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 485-502. Cold Spring Harbor Laboratory, New York.

Ishii, S., J.S. Salstrom, Y. Sugino, W. Szybalski and F. Imamato (1980) Gene 10, p. 17-25

Ito, K. (1972) Mol.Gen.Genet. 115, p. 349-353

Jackson, E.N. and C. Yanofsky (1973) J.Mol.Biol. 76, p. 89-101

Jacob, F., and J. Monod (1962) Cold Spring Harbor Symp.Quant.Biol. 26, p. 193-211

Jaurin, B., T. Grundström, T. Edlund and S. Normark (1981) Nature 290, p. 221-225

Johnston, P.D. and A.G. Redfield (1978) Nucleic Acids Res. 5, p. 3913-3927

Johnston, P.D. and A.G. Redfield (1981) Biochemistry 20, p. 1147-1156

Kano, Y., A. Matsushiro and Y. Shimura (1968) Mol.Gen.Genet. 102, p. 15-26

Kieffer, M., N. Neff and M.J. Chamberlin (1977) J.Virol. 22, p.548-552

Kolb, A. and H. Buc (1982) Nucleic Acids Res. 10, p. 473-485

Koller, T., O. Kübler, R. Portman and J.M. Sogo (1978) J.Mol.Biol. 120, p. 121-131

Korn, L.J. and C. Yanofsky (1976) J.Mol.Biol. 106, p. 231-241

Krüger, D.H. and C. Schroeder (1981) Microbiol.Rev. 45, p. 9-51

Kung, H.-F., C. Spears and H. Weissbach (1975) J.Biol.Chem. 250, p. 1556-1562

Küpper, H., T. Sekiya, M. Rosenberg, J. Egan and A. Landy (1978) Nature 272, p. 423-428

Ladner, J.E., A. Jack, J.D. Robertus, R.S. Brown, D. Rhodes, B.F.C. Clarke and A. Klug (1975) Proc.Natl.Acad.Sci.USA 72, p. 4414-4418

Lathe, R. (1978) Curr. Topics in Microbiol. and Immunol. 83, p. 37-91

Lauther, R.P. and G.W. Hatfield (1980) Proc.Natl.Acad.Sci.USA 77, p. 1862-1866

Lecocq, J.-P. and C. Dambly (1976) Mol.Gen.Genet. 145, p. 53-64

Lee, F., K. Bertrand, G. Bennett and C. Yanofsky (1978) J.Mol.Biol. 121, p. 193-217

Lee, F. and C. Yanofsky (1977) Proc.Natl.Acad.Sci.USA 74, p.4365-4369

Lomant, A.J. and J.R. Fresco (1975) Prog. Nucleic Acid Res. Mol. Biol. 15, p. 185-218

Losick, R. and M. Chamberlin (eds.) (1976) "RNA Polymerase". Cold Spring Harbor Laboratory, New York.

Lowe, P.A., D.A. Hager and R.R. Burgess (1979) Biochemistry 18, p. 1334-1352

Majors, J. (1975a) Proc.Natl.Acad.Sci.USA 72, p. 4394-4398

Majors, J. (1975b) Nature 356, p. 672-674

Maniatis, T., R.C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G.K. Sim and A. Efstratiadis (1978) Cell 15, p. 687-701

Maniatis, T., A. Jeffrey and J. van de Sande (1975) Biochemistry 14, p. 3787-3794

Martin, F. and I. Tinoco (1980) Nucleic Acids Res. 8, p. 2295-2300

McAllister, W.T. and C.L. Barrett (1977a) Virol. 82, p. 275-287

McAllister, W.T. and C.L. Barrett (1977b) J.Virol. 23, p. 543-553

McAllister, W.T. and H.-L. Wu (1978) Proc.Natl.Acad.Sci.USA 75, p. 804-808

McConnell, D. (1979a) Nucleic Acids Res. 6, p. 3491-3503

McConnell, D. (1979b) pers. comm.

McDermit, M., M. Pierce, D. Staley, M. Shimaji, M. Shaw and D.L. Wulff (1976) Genet. 82, p. 417-422

McDonell, M.W., M.N. Simon and F.W. Studier (1977) J.Mol.Biol. 110, p. 119-146

McKay, D.B. and T.A. Steitz (1981) Nature 290, p. 744-749

McKenney, K., H. Shimatake, D. Court, U. Schmeissner, C. Brady and M. Rosenberg (1981) "Gene Amplification and Analysis", vol. II p. 383-415. Pub. Elsevier-North Holland. Eds. J.G. Chirikjian and T.S. Papas

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

Merril, R.C., M.E. Gottesman and S.L. Adhya (1981) J.Bact 147, p. 875-887

Miller, J.H. (1972) "Experiments in Molecular Genetics". Cold Spring Harbor Laboratory, New York, p. 352-355

Millette, R.L., C.D. Trotter, P. Herrlich and M. Schweiger (1970) Cold Spring Harbor Symp.Quant.Biol. 35, p. 135-142

Minkley, E.G. and D. Pribnow (1973) J.Mol.Biol. 77, p. 255-277

Miozzara, G.F. and C. Yanofsky (1978) J.Bact. 133, p. 1457-1466

Morse, D.E. and A.N.C. Morse (1976) J.Mol.Biol. 103, p. 209-226

Morse, D.E. and P. Primakoff (1970) Nature 226, p. 28-31

Murray, K. (1973) Biochem.J. 131, p. 569-583

Murray, N.E., W.J. Brammar and K. Murray (1977) Mol.Gen.Genet. 150, p. 53-61

Musso, R.E., R. di Lauro, S. Adhya and B. de Crombrugghe (1977) Cell 12, p. 847-854

Newman, A.J. (1980) PhD Thesis. University of Edinburgh.

Newman, A.J., J.-C. Ma, K.M. Howe, I. Garner and R.S. Hayward (1982) submitted to Nucleic Acids Res.

Oakley, J.L. and J.E. Coleman (1977) Proc.Natl.Acad.Sci.USA 74, p. 4266-4270

Oda, T. and M. Takanami (1972) J.Mol.Biol. 71, p. 799-802

Ogden, S., D. Haggerty, C.M. Stoner, D. Kolodrubetz and R. Schleif (1980) Proc.Natl.Acad.Sci.USA 77, p. 3346-3350

O'Hare, K. (1978) PhD Thesis, University of Edinburgh

O'Hare, K. and R.S. Hayward (1981) Nucleic Acids Res. 9, p. 4689-4707

Oppenheim, D.O. and C. Yanofsky (1980) Genet. 95, p. 785-795

Ovchinnikov, Y.A., V.M. Lipkin, N.N. Modyanov, O.Y. Chertov and Y.V. Smirnov (1977) FEBS Lett. 76, p. 108-111

Ovchinnikov, Y.A., G.S. Monastyrskaya, V.V. Gubanov, S.O. Guryev, O.Y. Chertov, N.N. Modyanov, V.A. Grinkevich, I.A. Makarova, T.V. Marchenko, I.N. Polovnikova, V.M. Lipkin and E.D. Sverdlov (1981) Eur.J.Biochem. 116, p. 621-629

Ovchinnikov, Y.A., G.S. Monastyrskaya, V.V. Gubanov, S.O. Guryev, I.S. Salomatina, T.M. Shuvaeva, V.M. Lipkin and E.D. Sverdlov (1982) Nucleic Acids Res. 10, p. 4035-4044

Oxender, D.L., G. Zurawski and C. Yanofsky (1979) Proc.Natl.Acad. Sci.USA 76, p. 5524-5528

Panayotatos, N. and R.D. Wells (1979) Nature 280, p. 35-39

Peters, G.G. and R.S. Hayward (1974) Eur.J.Biochem. 48, p. 199-208

Pfennig-Yeh, M., H. Ponta, M. Hirsch-Kauffmann and H.-J. Rahmsdorf (1978) Mol.Gen.Genet. 166, p. 127-140

Platt, T. (1981) Cell 24, p. 10-23

Platt, T., C. Squires and C. Yanofsky (1976) J.Mol.Biol. 103, p. 411-420

Platt, T., K. Weber, D. Ganem and J.H. Miller (1972) Proc.Natl.Acad. Sci.USA 69, p. 897-901

Ponta, H., H.-J. Rahmsdorf, S.H. Pai, M. Hirsch-Kauffmann, P. Herrlich and M. Schweiger (1974) Mol.Gen.Genet. 134, p. 281-297

Post, L.E., G.D. Strychartz, M. Nomura, H. Lewis and P.P. Dennis (1979) Proc.Natl.Acad.Sci.USA 76, p. 1697-1701

Pribnow, D. (1975) Proc.Natl.Acad.Sci.USA 72, p. 784-788

Ptashne, M., A. Jeffrey, A.D. Johnson, R. Maurer, B.J. Meyer, C.O. Pabo, T.M. Roberts and R.T. Sauer (1980) Cell 19, p. 1-11

Queen, C. and M. Rosenberg (1981) Cell 25, p. 241-249

Quigley, G.J., N.C. Seeman, A.H.-J. Wang, F.L. Suddath and A. Rich (1975) Nucleic Acids Res. 2, p. 2329-2341

Ratner, D. (1976a) In "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 645-685. Cold Spring Harbor Laboratory, New York.

Ratner, D. (1976b) Nature 259, p. 151-153

Reid, B.R., L. McCollum, N. Ribeiro, J. Abbate and R.E. Hurd (1979) Biochemistry 18, p. 3996-4005

Richardson, J.P. (1970) Cold Spring Harbor Symp.Quant.Biol. 35, p. 127-133

Richardson, J.P. and R. Conaway (1980) Biochemistry 19, p. 4293-4299

Richardson, J.P., C. Grimley and C. Lowery (1975) Proc.Natl.Acad. Sci.USA 72, p. 1725-1728

Rigby, P.W.J., M. Dieckmann, C. Rhodes and P. Berg (1977) J.Mol.Biol. 113, p. 237-251

Roberts, J. (1969) Nature 224, p. 1168-1174

Roberts, J. (1976) In "RNA Polymerase", eds. R. Losick and M. Chamberlin, p. 247-271. Cold Spring Harbor Laboratory, New York.

Robertson, H.D., E. Dickson and J.J. Dunn (1977) Proc.Natl.Acad. Sci.USA 74, p. 822-826

Robillard, G.T., C.W. Hilbers, B.R. Reid, J. Gaugloff, G. Dirheimer and R.G. Schulman (1976) Biochemistry 15, p. 1883-1888

Romaniuk, P.J., D.W. Hughes, R.J. Grégoire, R.A. Bell and T. Neilson (1979) Biochemistry 18, p. 5109-5116

Rosa, M. (1979) Cell 16, p. 815-825

Rosenberg, M. and D. Court (1979) Ann.Rev.Genet. 13, p. 319-533

Rosenberg, M., D. Court, H. Shimatake, C. Brady and D.L. Wulff (1978) Nature 272, p. 414-423

Rosenberg, M. and R.A. Kramer (1977) Proc.Natl.Acad.Sci.USA 74, p. 984-988

Rosenberg, M., R.A. Kramer and J.A. Steitz (1974) J.Mol.Biol. 89, p. 777-782

Rosenberg, A.H., M.N. Simon, F.W. Studier and R.J. Roberts (1979) J.Mol.Biol. 135, p. 907-915

Saito, J., S. Tabor, F. Tamanoi and C.C. Richardson (1980) Proc. Natl.Acad.Sci.USA 77, p. 3917-3921

Salstrom, J.S. and W. Szybalski (1976) Fed.Proc. 35, p. 15-38

Salstrom, J.S. and W. Szybalski (1978a) Virol. 88, p. 252-262

Salstrom, J.S. and W. Szybalski (1978b) J.Mol.Biol. 124, p. 195-221

Sanger, F. and A.R. Coulson (1978) FEBS Lett. 87, p. 107-110

Sanger, F., S. Nicklen and A.R. Coulson (1977) Proc.Natl.Acad.Sci. USA 74, p. 5463-5467

Schaller, H., C. Gray and K. Herrmann (1975) Proc.Natl.Acad.Sci.USA 72, p. 737-741

Schmitz, A. and D. Galas (1979) Nucleic Acids Res. 6, p. 111-137

Schreier, P.H. and R. Cortese (1979) J.Mol.Biol. 129, p. 169-172

Segawa, T. and F. Imamoto (1976) Virol. 70, p. 181-184

Shigesada, K. and C.-W. Wu (1980) Nucleic Acids Res. 8, p. 3355-3369

Siebenlist, U. (1979) Nucleic Acids Res. 6, p. 1895-1907

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

Simpson, R. (1980) Nucleic Acids Res. 8, p. 759-766

Smith, A.J.H. (1974) Nucleic Acids Res. 6, p. 831-848

Southern, E.M. (1975) J.Mol.Biol. 98, p. 503-517

Squires, C.L. (1982) pers. comm.

Squires, C., F. Lee, K. Bertrand, C.L. Squires, M.J. Bronson and C. Yanofsky (1976) J.Mol.Biol. 103, p. 351-381

Stahl, S.J. and M.J. Chamberlin (1977) J.Mol.Biol. 112, p. 577-601

Stauffer, G.V., G. Zurawski and C. Yanofsky (1978) Proc.Natl.Acad. Sci.USA 75, p. 4833-4837

Steitz, J.A. and R.A. Bryan (1977) J.Mol.Biol. 114, p. 527-543

Studier, F.W. (1972) Science 176, p. 367-376

Studier, F.W. (1975) J.Mol.Biol. 94, p. 283-295

Studier, F.W. and J.V. Maizel, Jr. (1969) Virology 39, p. 575-586

Studier, F.W. and A.H. Rosenberg (1981) J.Mol.Biol. 153, p. 503-526

Studier, F.W., A.H. Rosenberg, M.N. Simon and J.J. Dunn (1979) J.Mol.Biol. 135, p. 917-937

Sutcliffe, J.G. (1978) Nucleic Acids Res. 5, p. 2721-2728

Takanami, M., K. Sugimoto, H. Sugisaki and T. Okamoto (1976) Nature 260, p. 297-302

Taniguchi, T., M. O'Neil and B. de Crombrugghe (1979) Proc.Natl. Acad.Sci.USA 76, p. 5090-5094

Taylor, K., Z. Hradecna and W. Szybalski (1967) Proc.Natl.Acad.Sci. USA 57, p. 1618-1625

Tessman, E.S. and P.K. Peterson (1976) J.Bacteriol. 128, p. 264-270

Tinoco, I., Jnr., P.N. Borer, B. Dengler, M.D. Levine, O.C. Uhlenbeck, D.M. Crothers and J. Gralla (1973) Nature New Biol. 246, p. 40-41

Travers, A.A. and R.R. Burgess (1969) Nature 222, p. 537-540

Uhlenbeck, O.C., J. Baller and P. Doty (1970) Nature 225, p. 509-510

Vogt, V. (1969) Nature 223, p. 854-855

Ward, D.F. and M. Gottesman (1981) Nature 292, p. 212-215

Ward, D.F. and N.E. Murray (1979) J.Mol.Biol. 133, p. 249-266

West, R.W., D. McConnell and R.L. Rodriguez (1980) Mol.Gen.Genet. 180, p. 439-447

Wozny, M.E., H.N. Carnevale and E.E. Jones (1975) Biochem.Biophys. Acta. 383, p. 106-116

Wu, A.M., G.E. Christie and T. Platt (1981) Proc.Natl.Acad.Sci.USA 78, p. 2913-2917

Wu, A.M. and T. Platt (1978) Proc.Natl.Acad.Sci.USA 75, p. 5442-5446

Wu, F.Y.-H., L.R. Yarbrough and C.-W. Wu (1976) Biochemistry 15, p. 3254-3258

Wulff, D.L. (1976) Genet. 82, p. 401-416

Yamamori, T., K. Ito, T. Yura, T. Suzuki and T. Iino (1977) J.Bact. 132, p. 254-261

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

Yanofsky, C. and L. Soll (1977) J.Mol.Biol. 113, p. 663-677

Zarudnaya, M.I., Y.N. Kosaganov, Y.S. Lazurkin, M.D. Frank-Kamenetskii, R.Sch. Beabealashvilli and L.P. Savochkina (1976) Eur.J.Biochem. 63, p. 607-615

Zillig, W., H. Fujiki, W. Blum, D. Janekovic, M. Schweiger, H.-J. Rahmsdorf, H. Ponta and M. Hirsch-Kauffmann (1975) Proc.Natl.Acad.Sci.USA 72, p. 2506-2510

Zillig, W., K. Zechel and H. Halbwachs (1970) Hoppe-Seyler's Z Physiol.Chem. 351, p. 221-224

Zurawski, G., D. Elseviers, G.V. Stauffer and C. Yanofsky (1978) Proc.Natl.Acad.Sci.USA 75, p. 5988-5992

Zurawski, G. and C. Yanofsky (1980) J.Mol.Biol. 142, p. 123-129

Zaruchi-Schultz, T., C. Jerez, G. Goldberg, H.-F. Kung, K.-H. Huang, N. Brot and H. Weissbach (1979) Proc.Natl.Acad.Sci.USA 76, p. 6115-6119