A characterisation of two genes of *Escherichia coli*, and the possible effect of BIMEs on gene expression.

STEPHEN WEDGWOOD

A thesis presented for the degree of Ph.D.

Institute of Cell and Molecular Biology
University of Edinburgh

September 1996

DECLARATION

I hereby declare that I alone have written this thesis and that, except where indicated, the work presented is my own.

Stephen Wedgwood **September 1996**

ACKNOWLEDGEMENTS

I wish to thank several people who have been of help to me during my period of study : First of all to Richard Hayward for the guidance and resourcefulness which has been of great help throughout, and also for the many social events he has organised. Many thanks to the rest of the group (especially Dave, Helen, Ashok, and Brenda) who have contributed towards a pleasant working atmosphere over the years. Thanks also to members of the Donachie/Masters groups who have offered welcome advice, and to Prof. Ishihama and others at N.I.G. for hospitality and assistance during my stay in Mishima. Cheers to various fellow partakers of the Waters of Life, especially Jules. Thanks to Laphroaig for inspiration, to the mighty Warriors of Stenhousemuir for proving that anything is possible, and to that fine gastrodome The Kebab Mahal. Finally I wish to thank my family for all their support throughout my many years in this excellent city.

ABBREVIATIONS

aa	amino acid(s)
AmpR	ampicillin resistant
AMPS	ammonium persulphate
bp	base pair(s)
cAMP	cyclic AMP
ccc	covalently closed circular
CTD	carboxy-terminal domain
BSA	bovine serum albumin
cDNA	complementary DNA
ddNTP	dideoxynucleoside 5′-triphosphate
dNTP	deoxynucleoside 5′-triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
dH_2O	distilled water
EDTA	diamino ethane tetraacetic acid
EtdBr	ethidium bromide
g	acceleration due to gravity
GST	glutathione S-transferase
HTH	helix-turn-helix
IPTG	isopropyl-β-D-thiogalactoside
KanR	kanamycin resistant
kb	kilobase
LB	Luria broth

moi	multiplicity of infection
MOPS	3-[N-morpholino] propane sulphonic acid
mRNA	messenger RNA
nt	nucleotide(s)
NTD	amino-terminal domain
PEG	polyethylene glycol
p.f.u.	plaque forming unit
ppGpp	guanosine 3′,5′-bis(pyrophosphate)
psi	pounds per square inch
RF	replicative form
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
ss	single-stranded
TEMED	N, N, N′, N′-tetramethyl ethylene diamine
Tris	tris (hydroxymethyl) aminomethane
v/v	volume to volume
w/v	weight to volume
X-Gal	5-bromo-4-chloro-3-inodyl-β-D-galactoside

ABSTRACT

In the course of work aimed at discovering genes encoding novel sigma subunits of *Escherichia coli* RNA polymerase, two unknown open reading frames were identified by hybridisation of an *E.coli* genome library with synthetic oligonucleotide probes directed against the conserved subregion 2.1 of bacterial sigma factors. This thesis describes the mapping, sequencing, and characterisation of these genes.

The open reading frame *f229* was located near 4324kb on the physical map of the *E.coli* K12 chromosome. Its sequence and that of a downstream Bacterial Interspersed Mosaic Element (BIME) were determined. Transcription of *f229* was not detectable under normal growth conditions, nor was it found to be inducible by various environmental shocks. However, *f229* mRNA levels increased dramatically following alteration of the BIME structure, either by complete removal or by a deletion leaving just a single Palindromic Unit. This transcription was shown to be σ^{70}-dependent, and the *f229* promoter was subsequently mapped. A protein encoded by *f229* was identified and had the expected molecular size of 25kDa, but was found to be poorly expressed even in the absence of the BIME. Moreover, an *f229::kanR* mutation did not affect growth of *E.coli* under a variety of conditions. Since the predicted amino acid sequence of F229 has no identity with any known protein and no similarity to members of either of the two known sigma families, the function of *f229* remains unknown.

The downstream portion of a second open reading frame was located near 381kb on the physical map. The amino acid sequence has very high identity (80% range) with a family of *Pseudomonas* enzymes, the 4-hydroxy-2-oxovalerate aldolases. These participate in a metabolic pathway that converts a variety of aromatic compounds into Krebs Cycle intermediates. Since *E.coli* is known to carry out similar catabolic reactions and the corresponding genes in *Pseudomonas* are clustered into operons, it is likely that the unsequenced region upstream of this open reading frame contains further genes involved in aromatic degradation.

CONTENTS

LIST OF FIGURES

Chapter 1

The aim of this project was to discover genes encoding novel sigma subunits of *E.coli* RNA polymerase. Sections 1.1 to 1.8 deal with *E.coli* RNA polymerase and its sigma subunits. One candidate sigma gene, *f229*, has a BIME structure at its 3′ end, which appears to influence *f229* transcription. BIMEs are discussed in section 1.9. Osmotic and cold-shock were thought to play a role in *f229* expression, and are discussed in sections 1.10 and 1.11 respectively. A second candidate sigma gene, *orf243*, is predicted to encode a protein having striking homology with the *Pseudomonas* 4-hydroxy-2-oxovalerate aldolases. These enzymes are involved in catabolism of aromatic compounds, which is discussed in section 1.12.

Chapter 1

Structure and function of eubacterial sigma factors

1.1 Introduction

Prokaryotic RNA polymerase (RNAP) consists of a core enzyme (subunits $\alpha_2\beta\beta'$) and one of several sigma (σ) subunits. The core enzyme is catalytically active for RNA polymerisation and termination, but the σ subunit is essential for promoter DNA sequence recognition (Lesley and Burgess, 1989). All eubacteria studied so far contain multiple σ species. To date, seven *Escherichia coli* sigmas have been identified, each conferring a distinct transcriptional specificity on RNAP by directing the enzyme to different promoter sequences. The major *E.coli* sigma is the product of the *rpoD* gene, σ^{70}, and is responsible for the majority of transcription, while at least six minor species transcribe co-ordinately regulated gene sets in response to various stress conditions (Table 1.1).

Table 1.1 *E.coli* sigmas

Sigma	Size (kDa)	Gene	Primary Function
70	70	*rpoD*	Major sigma factor
S	38	*rpoS*	Major stationary phase transcription
32	32	*rpoH*	Heat-shock response
E	24	*rpoE*	Response to extracytoplasmic protein damage
F	28	*rpoF*	Flagellar synthesis and chemotaxis
FecI	19	*fecI*	Citrate-dependent iron transport
54	54	*rpoN*	Nitrogen assimilation

Sigmas play a transitory role in transcription and are released shortly after initiation. They are generally present at a low level when compared stoichiometrically with core enzyme, and promote transcription in a catalytic manner; that is, one sigma allows many core polymerase molecules to initiate transcription (reviewed by Burgess and Travers, 1970).

Sequence analysis of many sigma factors has indicated that, generally speaking, they can be divided into two protein families, related in structure and function to σ^{70} or σ^{54}. As the initial aim of my project was the characterisation of genes which might encode novel sigmas of the σ^{70} family, I have concentrated below mostly on the structure and function of that family. A shorter discussion on the σ^{54} family is also included.

1.2 *E.coli* RNA polymerase: a brief overview

E.coli RNA polymerase is a large multisubunit DNA-dependent RNA polymerase consisting of two α, one β, one β', and one of a number of σ subunits (Table 1.2). The individual subunits show no biological activity when isolated, but instead must self-assemble in a defined order. This has been determined *in vivo* and *in vitro* by Ishihama and co-workers (reviewed by Ishihama, 1981) and can be summarised as follows:

$$\alpha + \alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \underset{\text{core enzyme}}{\alpha_2\beta\beta'} \rightarrow \underset{\text{holoenzyme}}{\alpha_2\beta\beta'\sigma}$$

Table 1.2 *E.coli* RNA polymerase major subunits

Subunit	Size (kDa)	Gene	Reference
α	36.5	*rpoA*	Post and Nomura (1979)
β	150.6	*rpoB*	Ovchinnikov *et al.* (1981)
β'	155.2	*rpoC*	Ovchinnikov *et al.* (1982)
σ	70.3	*rpoD*	Burton *et al.* (1981)

The core enzyme is capable of ribonucleotide polymerisation and termination at some sites, but is unable to initiate transcription at specific sites on double-stranded DNA (promoters). A functional sigma subunit is absolutely required to allow proper transcription initiation by holoenzyme (Burgess, 1969). Genetic and biochemical evidence has identified residues in the core subunits of *E.coli* RNA polymerase that are involved in transcriptional processes. These are briefly summarised below.

1.2.1 The α subunit

Dimerisation of the α subunit is the first step in the assembly pathway of RNAP, and thus α plays a key role in enzyme formation. A substantial portion of the C-terminus of α can be deleted without affecting its assembly into core, whilst point mutations within the N-terminal domain (α-NTD) block RNA polymerase assembly *in vitro* (Igarashi *et al.,* 1991) and *in vivo* (Hayward *et al.,* 1991). The minimum portion of α required for active enzyme assembly was narrowed down to a region from residues 21 to 235 (Kimura *et al.,* 1994). Insertion analysis (Kimura and Ishihama, 1995a) and amino acid substitution analysis (Kimura and Ishihama, 1995b) have identified two conserved sequences within this region which are involved in α dimerisation and in α-dimer interaction with the β and β′ subunits.

Similar studies using *rpoA* point mutants and C-terminally truncated α subunits have shown that a carboxy-terminal domain (α-CTD) is needed for transcription activation *in vitro.* This domain is thought to make protein-protein contacts with class I transcriptional activators which bind DNA sequences upstream of the -35 promoter element (reviewed by Ishihama, 1993), and protein-DNA contacts with UP (upstream activation) elements (Ross *et al.,* 1993; Gaal *et al.,* 1996). UP elements are A:T-rich sequences of ~20bp located immediately upstream of the -35 element of strong promoters such as *rrnB* P1, and interaction with α-CTD can increase the efficiency of initiation by some 30-fold. α-NTD and α-CTD are connected by a flexible interdomain linker which is thought to allow the α-CTD to make different interactions at different promoters (Blatter *et al.,* 1994).

Recently Liu *et al.* (1996) reported an additional role for α in transcription pausing, termination, and anti-termination involving a direct interaction between α and the elongation factor NusA. The α-CTD was found to stabilise the interaction between NusA and the transcription complex, allowing efficient modulation of elongation, termination and anti-termination.

1.2.2 The β and β′ subunits

The β and β′ subunits are highly conserved throughout prokaryotic and eukaryotic RNA polymerases with regions of high sequence similarity separated by regions that are poorly conserved (Rowland and Glass, 1990). The active centre of RNAP, which performs the principal biochemical reaction of gene expression, appears to be comprised of distinct segments within the core enzyme. Crosslinking studies of ternary complexes have identified two regions in β close to the 5′ initiating nucleotide (Mustaev *et al.*, 1991) and one region in β′ near the 3′ end of the nascent RNA chain (Borukov *et al.*, 1991) implicating these regions in the formation of the active site. Phenotypes of substitution mutations in the immediate vicinity of these crosslinks in the β subunit are consistent with the idea that these regions of the β polypeptide participate in forming the catalytic centre of the enzyme (Mustaev *et al.*, 1991; Sagitov *et al.*, 1993).

Analysis of the effect of defined amino acid (aa) substitutions on RNAP function *in* vivo located sites on the β polypeptide involved in the resistance to rifampicin, in transcription termination, and in the stringent response (Nene and Glass, 1984). The β subunit is the target site for the transcription inhibitors rifampicin and streptolydigin, which respectively abort initiation and block elongation (Oen and Wu, 1978; Iwakuri *et al.*, 1973). Most mutations conferring rifampicin resistance map to a central region of β (Jin and Gross, 1988; Severinov *et al.*, 1993), although other mutations have been located at the N-terminus (Severinov *et al.*, 1994). Mutations conferring streptolydigin resistance were originally found exclusively in

the β subunit (Heisler *et al.,* 1993) but more recently a mutation was found in β′ (Severinov *et al.,* 1995).

The β subunit is a target for ppGpp, which mediates the inhibition of stable RNA synthesis during the so-called stringent response (see section 1.5.7). Single aa changes in β were found to render RNAP resistant to ppGpp, with the nature of the promoter also playing a part in the strength of the stringent control signal (Glass *et al.*, 1986a and 1987).

β and β′ are also involved in RNA elongation and termination, with amino acid substitutions in both subunits affecting these properties (Weilbacher *et al.,* 1994; Heisler *et al.,* 1996 and references therein). Tavorima *et al.* (1996) also provided evidence of intramolecular interaction between conserved regions of β. They identified suppressors of previously identified lethal β mutations; some of these suppressor mutations were themselves, when isolated, unable to support cell growth. The accumulating evidence suggests that functional units within RNAP may not necessarily be confined to one particular subunit. Contributions may come from segments of polypeptides that are not contiguous in their primary sequence, and also from segments of other subunits. However, there is still the possibility that some of the regions identified are indirectly rather than directly involved in the functions ascribed. The role of sigma is discussed in detail below.

1.3 The σ^{70} family: functions in brief

Over 37 members of the σ^{70} family have been isolated and sequenced from a variety of eubacteria (Lonetto *et al.,*1992 and 1994). All promoters recognised by holoenzymes containing sigmas of this type consist of blocks of conserved sequences which are located approximately 10 and 35 bp upstream of the transcription start site. Conserved promoter sequences recognised by different forms of holoenzyme can vary in the identity, length and spacing between the -10 and -35 elements (Doi and Wang, 1986, and table 1.3). It is widely accepted that members of

the σ^{70} family are sequence-specific DNA binding proteins, and sigma is the only subunit of RNAP holoenzyme that can be crosslinked to both the -10 and -35 promoter regions (Chenchick *et al.,* 1981). Promoter recognition is discussed in sections 1.5.6 and 1.5.8.

Following promoter recognition, sigmas play an essential role in the subsequent steps of transcription initiation, in particular the localised melting (DNA strand separation) around the start site. Structural and kinetic studies have shown that the opening of DNA strands at both σ^{70} and σ^{32}-dependent promoters involves at least two intermediate RNAP-DNA complexes which show different protection patterns in DNaseI, hydroxyl radical, and permanganate footprinting (reviewed by Gross *et al.,* 1992; deHaseth and Helmann, 1995). It appears that initial binding of the promoter by RNAP holoenzyme results in the formation of a closed complex which isomerises, via a second intermediate, to the open complex in which the DNA is unwound between -11 and +3 relative to the transcription start site at +1 (Chan *et al.,* 1990). Possible roles for regions of sigma in open complex formation are discussed in section 1.5.5.

Table 1.3 Promoter sequences recognised by *E.coli* sigmas

Sigma	-35 element	-10 element	Reference
70	TTGACA	TATAAT	Hawley and McClure (1983)
S	?	TATAAT	Tanaka *et al.* (1995)
32	TCTC-CCTTGAA	CCCCAT-TA	Gross *et al.* (1990)
E	GAACCT	TCTGA	Erickson and Gross (1989)
F	CTAAA	CCGATAT	Helmann (1991)
FecI	AAGGAAAAT	TCCTTT	Enz *et al.* (1995)
54	CTGG-A	TTGCA	Gussin *et al.* (1986)

1.4.1 Group 1 sigmas

Using pairwise sequence alignments and functional criteria, Lonetto *et al.* (1992) divided the σ^{70} family into three groups. Group 1 sigmas are the primary sigmas of diverse organisms, and are responsible for the majority of gene transcription during exponential growth. They are highly homologous, sharing pairwise sequence identities of at least 51%. Three regions of amino acid sequence are almost invariant among primary sigmas, and can be used to distinguish them from group 3 sigmas (section 1.4.3). Two of these regions, the "*rpoD* box" (Tanaka *et al.*, 1988) and a conserved 20aa sequence, overlap the -10 and -35 promoter recognition determinants. This emphasises the close functional relationship of the primary sigmas, which recognise closely related promoter sequences. Indeed, the major sigma of *Bacillus subtilis* (σ^{43}) and *E.coli* σ^{70} direct polymerase to identical consensus DNA sequences and can function *in vitro* with each other's core (Shorenstein and Losick, 1973). This is despite a major difference in size, due mainly to an extra 245aa between regions 1.2 and 2.1 of σ^{70}. Recently, these aa have been shown by deletion to be dispensible for the basic functions of *E.coli* σ^{70} *in vivo* (Kumar *et al.*, 1995).

1.4.2 Group 2 sigmas

Group 2 sigmas include *E.coli* σ^{S} (section 1.8.1.) and three alternative sigmas of *Streptomyces coelicolor* (HrdA, HrdC, and HrdD) and are non-essential for exponential cell growth. They show strong homology to group 1 sigmas, especially in their DNA binding regions, which may reflect overlapping promoter specificities. Tanaka *et al.* (1993) demonstrated that *E.coli* σ^{70} and σ^{S} recognise a common subset of *E.coli* promoters, whilst each was found to display a distinct specificity for promoters not recognised by the other. σ^{S} is discussed in more detail in section 1.8.1.

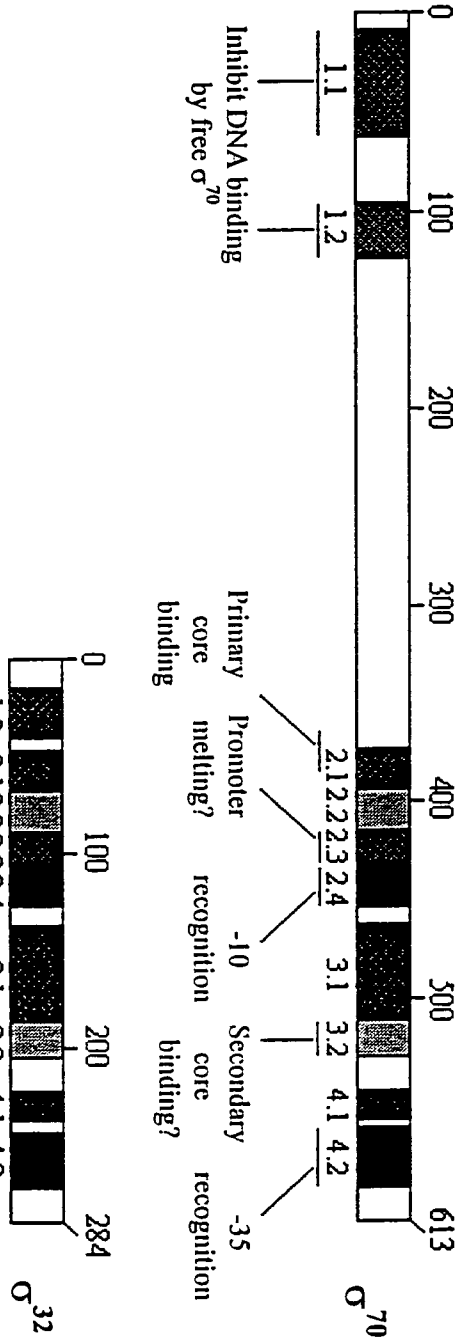
1.4.3 Group 3 sigmas

Group 3 sigmas are alternate sigmas which show significant deviation from the primary sigma sequences, exhibiting a maximum identity of 27%. The group 3 sigmas tend to cluster into functional subgroups that include members from diverse organisms. Members of a subgroup are more similar than primary and alternate sigmas in the same organism, suggesting that sigmas which regulate similar activities have common functional constraints. Accordingly, the heat-shock sigmas of *E.coli* and *Citrobacter freundii* share ~94% identity, whilst *E.coli* σ^{70} and σ^{32} are only 24% identical. The more recent identification of the Extracellular Function (ECF) family of sigma factors has considerably increased the overall diversity of the σ^{70} family (Lonetto *et al.,* 1994).

1.5 Sequence conservation of the σ^{70} family

Sequence alignments of the σ^{70} family members from several eubacteria have indicated that they are homologous proteins with four regions of high conservation (Helmann and Chamberlin, 1988; Lonetto *et al.,* 1992; fig. 1.1). Regions 2 and 4 are the most highly conserved and tend to be very basic, which is consistent with their postulated role in DNA binding. Regions 1 and 3 exhibit lower conservation and are acidic. The roles of the individual regions are discussed below.

Figure 1.1. Conserved sequences in the σ^{70} family (after Lonetto *et al.*, 1992). σ^{70} and σ^{32} are representative of major and alternate sigmas respectively. The aa scales are shown above each protein. Shaded rectangles represent conserved regions which are identified below. Known or proposed regional functions are indicated below σ^{70}.



1.5.1 Region 1

Region 1 is divided into two subregions, 1.1 and 1.2. The former is present only in the primary sigmas and *E.coli* σ^S whilst subregion 1.2 is present (to varying degrees) within all primary and almost all alternative sigmas, an exception being *S.typhimurium* FliA (Lonetto *et al.*, 1992). Although free σ^{70} is unable to bind DNA, Dombroski *et al.* (1992) reported that deletion of subregion 1.1 allowed non-specific DNA binding by the N-terminally-truncated sigma fused to GST. Moreover, a greater truncation affecting also subregion 1.2 and sequences downstream allowed σ^{70} to specifically bind promoter DNA. The obvious interpretation of these results is that subregion 1.1 sequesters the non-specific affinity of free σ^{70} for DNA whilst amino acids further donwnstream suppress promoter recognition, either alone or in conjunction with subregion 1.2. Gopal *et al.* (1994) provided evidence for specific interaction between a tryptophan residue at position 434, in the junction of 2.3 and 2.4, and the hydrophobic subregion 1.1. Dombroski *et al.* (1992) proposed that the N-terminus of σ^{70} normally masks the promoter recognition determinants which are only exposed upon interaction with core enzyme. According to their proposal, region 1 would become buried within the core in order to free the DNA binding regions of sigma.

Although *E.coli* σ^{32} lacks subregion 1.1, it too was unable to bind DNA in the absence of core (Dombroski *et al.*, 1993). However, deletion of the first 46aa including a non-conserved 18aa sequence allowed DNA binding, suggesting that σ^{32} contains its own amino-terminus inhibitory domain. On the other hand, the *S.typhimurium* sigma factor FliA, which lacks both subregions 1.1 and 1.2, was able to bind promoter DNA in the absence of core (Dombroski *et al.*, 1993). These results suggest that the interaction between different sigma factors and DNA can be regulated at the level of protein structure and/or conformation, with the amino terminal extensions providing the means for control of binding.

1.5.2 Region 2

Region 2 is highly conserved throughout eubacterial sigmas and is divided into four subregions (2.1-2.4) which have been implicated in core binding, open complex formation, and -10 promoter element recognition.

1.5.3 Subregion 2.1

Aa 361-390 of σ^{70}, constituting subregion 2.1 and an amino acid sequence immediately upstream, have been suggested to represent an essential minimal core-binding region by Lesley and Burgess (1989). Using a set of internal deletions and frameshift mutations within *rpoD* to determine the regions of sigma inessential for core binding, their results implicated this segment as containing the primary core-binding determinants. However, part of this region (aa 361-374) is located within the non-conserved 245aa insertion (characteristic in position, although variable in size, in Gram negative bacterial primary sigmas) suggesting that the essential core binding region can be further defined to the region immediately downstream of aa 374. It should be noted that this 245aa insertion can be deleted from σ^{70} without affecting its basic functions *in vivo* (Kumar *et al.,* 1995; section 1.4.1).

The β subunit is involved in the binding of σ to core. Although the removal of up to 20% of the C-terminus of β did not prevent the formation of the core enzyme, it was found to interfere with holoenzyme production (Glass *et al.*, 1986b).

1.5.4 Subregion 2.2

Subregion 2.2 has a highly conserved sequence in all groups of sigmas, with four invariant and several conserved residues, and was previously proposed as a possible core binding region (Helmann and Chamberlin, 1988). However, this subregion can be deleted without affecting the core-binding activity of σ^{70} (Lesley and Burgess,1989), and at present has not been assigned any specific function.

1.5.5 Subregion 2.3

Helmann and Chamberlin (1988) proposed that subregion 2.3, in conjunction with 2.1, might participate in DNA strand separation during open complex formation. This prediction was based upon the prevalence of evolutionarily conserved aromatic and basic amino acids in these subregions, in patterns which are reminiscent of the single-stranded DNA binding proteins of *E.coli* and its bacteriophages fd and T4. Part of subregion 2.3 is also similar to the "RNP-1 consensus sequence" shared by several eukaryotic proteins that bind to single-stranded RNA. Structural analysis of RNA and single-stranded DNA binding proteins has shown that aromatic amino acids facilitate nucleic acid -protein interactions by stacking between the nucleotide bases (Ollis and White, 1987). In addition, subregion 2.3 is in close proximity to amino acids that contact part of the -10 promoter element (section 1.5.6). Since σ^{70} can be crosslinked with the non-template strand at open complexes (Simpson *et al.*, 1979; Buckle *et al.*, 1991), it is plausible that the sigma subunit is involved in DNA melting around the transcription start site.

Juang and Helmann (1994) identified amino acid substitutions in subregion 2.3 of σ^{43} of *B.subtilis* (which is interchangable with σ^{70} of *E.coli*-section 1.4.1) that had significant effects on open complex formation. Some of these mutants also exhibited promoter-specific effects, suggesting that downstream -10 sequence recognition may be carried out by subregion 2.3 in a manner associated with promoter melting. In contrast, Waldburger and Susskind (1994) showed that most residues in σ^{70} subregion 2.3, including the aromatics and other residues showing similarity to the RNP-1 motif, were in fact tolerant of amino acid substitutions. Only one residue, Thr429, was intolerant to substitution suggesting an important role for this residue in σ^{70} function. Thr429 corresponds to a position that is not conserved in the RNP-1 motif. These discrepancies may be partly due to the different conditions used, with Juang and Helmann employing an *in vitro* assay system and temperature manipulation to increase the stringency of their tests.

It has been proposed that during open complex formation, the σ subunit binds the displaced non-template strand, rendering the latter insensitive to permanganate ions. Meanwhile, the template strand associates with the catalytic site located on β/β', suggested by the observation that a cold-sensitive β mutant has defects in promoter melting (Larionov *et al.,* 1979). Clearly, further experiments are required to fully understand the role of subregion 2.3 *in vivo* and *in vitro* during open complex formation.

1.5.6 Subregion 2.4

Subregion 2.4 contains three conserved hydrophobic residues located at every fourth amino acid, followed by conserved basic residues, and is predicted to have an alpha-helical secondary structure (Helmann and Chamberlin, 1988). Identification of amino acid substitutions in subregion 2.4 that suppress point mutations in -10 promoter regions indicate that the exposed face of the putative α-helix is involved in the recognition of the -10 promoter element.

Two amino acid substitutions affecting residue 437 in subregion 2.4 of σ^{70} have been isolated by selection for such suppressors. Q437H which substitutes the glutamine residue for a histidine suppressed a T to C mutation in the first position of the -10 hexamer (CATAAT) permitting a ~10-fold increase in the level of transcription from the mutant *lac*UV5 promoter relative to wild-type σ^{70} (Waldburger *et al.,* 1990). A second substitution mutant, T440I (Siegele *et al.,* 1989) was found to suppress C or G mutations at the first position of the -10 hexamer by 2- to 3-fold. Both of the above mutations are allele specific, that is, no other mutations in the -10 or -35 promoter regions were suppressed. Whilst T440I is a "loss of specificity" mutation allowing recognition of both consensus and mutant promoters, it is not clear whether Q437H alters the specificity of σ^{70} or results in loss of specificity for -10, since the mutant protein cannot be assayed *in vivo* in the absence of wild-type σ^{70}.

Similar studies have been used to pinpoint residues in subregion 2.4 of the *B.subtilis* alternative sigma σ^H, which affect recognition of the first two base pairs of the -10 element. T100I suppressed a G to A mutation at the first position of the -10 (Zuber *et al.*, 1989; Daniels *et al.*, 1990) whilst R96A suppressed an A to G substitution at the second position (Daniels *et al.*, 1990). These two mutations also cause allele-specific changes of specificity. The results suggest that residues located four amino acids apart govern the recognition of adjacent promoter base pairs, providing strong support for the involvement of an α-helix. σ^{70} and σ^H recognise different promoter sequences and show poor conservation in subregion 2.4. However, when the amino acid sequences are aligned residues implicated in contacting the first position of the -10 helix are adjacent in both proteins, suggesting that both primary and alternative sigmas may recognise promoters in a similar manner (Waldburger *et al.*, 1990).

Dombroski *et al.* (1992) showed that a GST fusion protein containing regions 2 and 3 of σ^{70} selectively recognised the -10 consensus sequence. This provides further evidence for the involvement of subregion 2.4 (and possibly 2.3, along with 2.2) in -10 promoter element recognition. Recently Minchin, Busby, and colleagues (personal communication) have identified a single aa alteration just downstream of subregion 2.4 (a region which they have named 2.5) which suppresses a down mutation in the TG element of an "extended -10" promoter (section 1.6). Possibly 2.4 and 2.5 really constitute a continuous DNA-recognition element.

1.5.7 Region 3

Region 3 is divided into subregions 3.1 and 3.2 (Lonetto *et al.*, 1992). Subregion 3.1 shows weak similarity with a helix-turn-helix DNA binding motif, although no function has previously been assigned for this subregion. However, recent work by Hernandez and Cashel (1995) demonstrated that spontaneous mutations in subregion 3.1 of σ^{70} suppressed physiological defects displayed by strains completely lacking guanosine 3′,5′-bis(pyrophosphate), ppGpp.

Following nutritional limitation of enteric bacteria there is a rapid inhibition of stable tRNA and rRNA synthesis, the so-called stringent response, mediated by a massive accumulation of ppGpp. The negative control of gene expression is accompanied by an induction of certain classes of mRNA such as amino acid biosynthesis transcripts (reviewed by Cashel and Rudd, 1987). Strains unable to synthesise ppGpp, a condition designated as ppGpp°, cannot grow in minimal media in the absence of amino acid supplements (Xiao *et al.,* 1991). The only mutations identified so far that suppress this phenotype occur in *rpoBC* (97%) and *rpoD* (3%) (Cashel *et al.,* 1995). The mechanism of the control of gene expression by ppGpp is unknown, although the suppressor mutations identified suggest that effects on promoter recognition, open complex formation and/or transcription elongation may be involved. The *rpoD* mutations characterised mapped to 3.1, suggesting a role for this subregion in ppGpp-mediated gene regulation. The synthesis of σ^S early in stationary phase is positively regulated by ppGpp and the accumulation of this sigma is deficient in a ppGpp° strain (Gentry *et al.,* 1993; section 1.8.1). Interestingly, Hernandez and Cashel (1995) found an increase in σ^S upon entry into stationary phase in ppGpp° strains carrying the subregion 3.1 mutations. Further *in vitro* studies involving ppGpp/RNAP interaction are underway.

Subregion 3.2 has several acidic amino acids which are conserved throughout the σ^{70} family. It may play a secondary role in core binding in some sigmas since a 25aa deletion affecting subregion 3.2 of σ^{32} reduced its affinity for core enzyme (Zhou *et al.*, 1992). However, many group 3 sigmas lack subregion 3.2 and it can be deleted from σ^{70} without significantly affecting core binding (Lesley and Burgess, 1989). It thus seems unlikely that 3.2 represents an essential core binding region. A possible role for 3.2 in contacting certain class II transcriptional activators has been suggested; this is discussed in section 1.5.8.

1.5.8 Region 4

Region 4 has been divided into two subregions (4.1 and 4.2) which are separated by a spacer region of variable length and amino acid sequence. Secondary structure predictions and sequence comparisons have led to the proposal that subregion 4.1 is an amphipathic α-helix, whilst 4.2 forms a classic helix-turn-helix (HTH) motif which is known in many cases to bind double-stranded DNA in a sequence-specific manner (Helmann and Chamberlin, 1988). Several lines of evidence indicate that subregion 4.2 recognises the -35 promoter element.

Allele-specific suppressor mutations in the proposed recognition helix of region 4.2 of σ^{70} were isolated, which compensated for down mutations at the third and fifth positions of the -35 consensus (TTGACA). An R584C substitution affecting the first amino acid of the proposed recognition helix suppressed the effects of a G or T substitution for C at position 5 of the -35 element (Siegele *et al.,* 1989). Similarly, substitutions of A or T for G at position 3 of -35 were suppressed by an R588H mutation of the fifth amino acid in the recognition helix (Gardella *et al.,* 1989). Further evidence of a role for 4.2 in -35 promoter recognition came from mutations at position 233 within subregion 4.2 of *B.subtilis* σ^{F}. An alanine to valine substitution changed its promoter specificity to that recognised by σ^{G}, whilst a substitution of methionine allowed σ^{F} to recognise σ^{B}-dependent promoters (Margolis *et al.,* 1991).

Siegele *et al.* (1989) isolated an additional set of mutants affecting amino acids in the structural helix of subregion 4.2 and the amino acid sequence immediately upstream of it, which non-specifically suppressed a series of promoter mutations. Based on structural and mutational studies carried out on other HTH proteins, it is likely that these σ^{70} mutations allow additional contacts with the sugar-phosphate backbone of the DNA in the vicinity of the -35 promoter element.

Region 4.2 has also been implicated in contacting class II transcriptional activator proteins which bind to promoter DNA in a position at or near the -35 region. The corresponding promoters often lack a consensus -35 sequence, and unlike class I

activators, transcription does not require the contact site on α (Ishihama, 1993), suggesting that protein-protein contacts with the carboxy-terminus of σ^{70} may be involved. Using a set of deletion mutants of σ^{70} truncated to different extents at the carboxy-terminus, Kumar *et al.* (1994) mapped the contact region of σ^{70} required to give activation of *in vitro* transcription initiation from appropriate promoters using the activators PhoB and CRP-cAMP. Whilst CRP-cAMP interaction required a segment extending from within subregion 3.2 to a region upstream of subregion 4.1, PhoB interaction required the first helix of subregion 4.2, consistent with the point mutation studies of Makino *et al.* (1993). A series of 4.2 point mutations generated by Kim *et al.* (1995) provided further evidence that the first helix is required for the interaction between RNAP and PhoB, but not for interaction with CRP-cAMP. More recently Makino *et al.* (1996) identified the DNA-binding domain in its C-terminus and four amino acids were found to be important for contacting RNAP. It appears that necessary contact and/or activation sites for class II activators lie at different positions in a segment of σ^{70} extending from subregion 3.2 to beyond subregion 4.2. It is plausible that other class II activators require contacts within 4.2 in order to activate transcription.

Dombroski *et al.* (1992) found that a GST fusion with region 4 of σ^{70} selectively recognised the -35 consensus sequence, whilst a fusion containing regions 2 and 4 selectively recognised both promoter elements. Although these results provide further evidence of roles for subregions 2.4 and 4.2 in promoter recognition, contacts with only 3 of 12bp defining a σ^{70} promoter are as yet understood. Additional amino acid contacts may be elucidated when the physical structure of sigma and its complex with DNA and the RNAP core are known.

Subregion 4.1 may act as a supporting structure for the HTH motif of subregion 4.2, based on structural similarity with bacteriophage λ Cro repressor, and with homeodomain proteins which also contain α-helices upstream of the HTH motif (Lonetto *et al.*, 1992).

1.6 Hybrid and truncated sigmas

In view of the high degree of sequence and functional conservation amongst members of the σ^{70} family, and the fact that all act in combination with the same species of core enzyme, it seemed plausible that a functional hybrid sigma (with regions 2.4 and 4.2 derived from different proteins) should be able to direct RNA polymerase to initiate from correspondingly hybrid promoters. To test this theory, Hayward *et al.* (1992) constructed a hybrid *rpoD-rpoH* gene encoding a hybrid sigma, σ^{7032}. This protein consists of amino acids 1-529 of σ^{70}, including subregion 2.4, and the terminal 82 amino acids of σ^{32}, including subregion 4.2. Whilst the hybrid sigma was unable to initiate transcription at σ^{32}-dependent promoters, it allowed vigorous expression *in vivo* of a reporter gene from a correspondingly hybrid promoter consisting of a -35 heat-shock element suitably spaced from a -10 major promoter element (Hayward *et al.*, 1992; Kumar *et al.*, 1995). The hybrid sigma may be more suitable for studies on structure/function relationships than either of its "parents", since work is difficult with σ^{70} due to the indispensability of *rpoD* and with σ^{32} due to its instability and the complex regulation of *rpoH*.

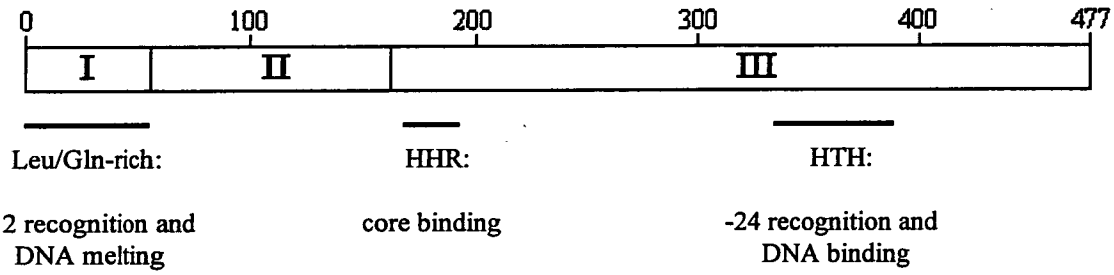
The positively regulated λP_{RE} promoter shows negligible (1/6) fit to the consensus -35 sequence, and only a poor (3/6) fit at -10. The promoter is barely recognised by purified RNAP and transcription is highly dependent on the class I activator protein cII (Shimatake and Rosenberg, 1981). In an effort to construct a constitutive derivative of the λP_{RE} promoter Keilty and Rosenberg (1987) perfected the -10 element to fit the consensus. The resulting promoter, P_{RE}*, allowed efficient cII-independent initiation of transcription *in vivo* and *in vitro*. Point mutations within P_{RE}* showed that while changes in the -35 element had only minor effects on promoter strength, sequences immediately upstream of the -10 consensus were critical for -35-independent activity. Keilty and Rosenberg concluded that the "extended -10" sequence TGnTATAAT constituted a -35-independent promoter and that the TG motif was an important requirement for efficient initiation at such promoters.

Since the hybrid sigma work described above suggested that regions 2.4 and 4.2 function as effectively independent promoter recognition domains, it was hypothesised that a carboxy-truncated σ^{70} lacking region 4 might retain the ability to allow efficient initiation at an extended -10 promoter. Results showed that both intact and truncated σ^{70} allowed purified RNAP to initiate efficiently and specifically at an extended -10 promoter, whereas only the intact σ^{70} permitted initiation at normal promoters lacking the TG motif (Kumar *et al.,* 1993). Whilst the extended -10 promoter is rarely if ever used in *E.coli* in the absence of a -35 element, it is probably used frequently to strengthen promoters.

1.7 The σ^{54} family

Around 20 homologues of σ^{54} have been identified, and studies have demonstrated that this sigma is both structurally and functionally distinct from the σ^{70} family. Initially identified as a positive regulator of nitrogen assimilation, σ^{54} is now known to be required for the expression of a wide variety of genes involved in diverse functions including dicarboxylic acid transport, catabolism of aromatic compounds, hydrogenase biosynthesis, and pilus production (reviewed by Kustu *et al.,* 1989). 17 fully characterised *rpoN* genes show a high degree of conservation in the proteins primary sequence and all conform to the three domain structure first proposed by Merrick *et al.* (1987). A poorly conserved acidic domain (region II) of variable length (25-100 residues) is flanked by an N-terminal glutamine and leucine-rich domain (region I) and a highly conserved C-terminal region (region III) ~400aa long (fig. 1.2). Region I is involved in -12 promoter element recognition and DNA melting (Hsieh and Gralla, 1994). Region II contains residues essential for binding to core RNA polymerase whilst region III, which carries a putative HTH motif, is involved in -24 promoter element recognition (Wong *et al.,* 1994). Region III also contains an almost totally conserved run of 10aa which has been called the "RpoN box". Despite its remarkable conservation a number of residues in the RpoN box can

Figure 1.2. Conserved sequences in the σ^{54} family. Regions are indicated within their corresponding rectangle. Motifs are identified below the protein along with their known or proposed functions. Leu/Gln rich is a Leucine/Glutamine-rich region; HHR is a hydrophobic heptad repeat; HTH is a helix-turn-helix motif.



be altered without inactivating the protein and its role remains to be determined (reviewed by Merrick, 1993).

σ^{54}-dependent promoters are markedly different in structure from those recognised by the σ^{70} family, comprising elements containing highly conserved GC and GG doublets centred near -12 and -24, respectively, relative to the transcription start site (table 1.3). σ^{54} alone is capable of specific binding to certain promoters, although its binding affinity is greatly enhanced in the presence of core enzyme (Buck and Cannon, 1992). The Eσ^{54} holoenzyme can form a closed promoter complex but is incapable of proceeding to an open complex in the absence of an activator protein (Sasse-Dwight and Gralla, 1988). These bind to specific DNA sites and normally bind within 100bp upstream of the promoter, but activation can occur with the activator located in excess of 1kb upstream. (Reitzer and Magasanik, 1986). In contrast to the case for class I activators at σ^{70}-dependent promoters (section 1.2.1) the α subunit C-terminus of the holoenzyme is not required for activation of initiation by Eσ^{54} (Lee *et al.,* 1993). The activator catalyses strand separation by a mechanism requiring ATP hydrolysis (Sasse-Dwight and Gralla, 1988) and once formed, the open complex can be maintained without the activator present (Popham *et al.,* 1989). *E.coli* σ^{54} is discussed in more detail in section 1.8.6.

1.8 *E.coli* alternative sigmas

From the sequence data available, it has been estimated that the *E.coli* chromosome carries 4000 genes, among which at least 1000 are expressed under normal laboratory conditions, i.e. at 37°C and with aeration (Helmann and Chamberlin, 1988). The rest of the genes are considered to be expressed under various stress conditions that *E.coli* meets in nature (Magasanik, 1982; Neidhardt *et al.,* 1984; Kolter *et al.,* 1993; Hengge-Aronis, 1993). Control of expression of coordinately regulated gene sets required for adaptation to various environmental conditions is often mediated by alternative sigmas, of which *E.coli* has at least 6.

1.8.1 Sigma S

In *E.coli,* it is known that expression of a set of genes is induced when cells enter the stationary phase and/or encounter stressful conditions for growth, such as nutrient starvation. Under such conditions, *E.coli* develops a general resistance against various physical and chemical stresses including heat, oxidative agents and osmotic shock (Hengge-Aronis, 1993; Kolter *et al.,* 1993). In addition, the cells metamorphose into a rather spherical shape and become smaller. These physiological and morphological changes associated with stationary growth phase have been studied at the molecular level (Hengge-Aronis, 1993; Kolter *et al.,* 1993; Siegele and Kolter, 1992). The global change in gene expression seems to be accompanied by modulation of the transcription and translation apparatus. For example, a high proportion of ribosomes are converted into dimers through the action of ribosome modulation factor (Wada *et al.,* 1995), and the major sigma factor σ^{70} is replaced by σ^{S}, the product of *rpoS* (Hengge-Aronis, 1993). In addition, RNA polymerase core is modified so as to change its promoter preference (Ozaki *et al.,* 1991; Ozaki *et al.,* 1992).

At least 30 genes are estimated to show *rpoS*-dependent expression (McCann *et al.,* 1991). *rpoS* was originally identified as *katF*, a positive regulator of *katE*, which encodes a stationary phase-specific catalase, HPII (Mulvey and Loewen, 1989). The amino acid sequence, deduced from the *katF* DNA sequence, is very similar to that of σ^{70} in regions 2 and 4, which include the specific DNA binding regions, making it a group 2 sigma (Lonetto *et al.,* 1992). That σ^{S} is indeed an alternative sigma was demonstrated by Tanaka *et al.* (1993). Using reconstituted RNA polymerase they showed that holoenzyme containing the *katF/rpoS* product, but not σ^{70}, was able to efficiently trigger *in vitro* transcription from a small number of appropriate promoters (although it was noted that a great many promoters can be recognised by polymerase carrying either σ^{70} or σ^{S}). Knockout mutations in *rpoS* reduce resistance to starvation, hydrogen peroxide and thermal stress, and drastically alter the pattern of protein synthesis upon starvation (McCann *et al.,* 1991).

Control of σ^S function

Since σ^S functions predominantly under certain growth conditions, its activity and/or cellular concentration have to be tightly controlled, although the mechanisms involved are unclear. The cellular content of σ^S increases ~7-fold upon entry into stationary phase (Tanaka *et al.*, 1993) and ~8-fold following carbon starvation (McCann *et al.*, 1993). Control of the cellular level of σ^S during starvation and upon entry into stationary phase appears largely posttranscriptional, involving stimulation of translation as well as changes in σ^S stability (Lange and Hengge-Aronis, 1994). These authors reported a mRNA secondary structure similarity between *rpoS* and *rpoH* transcripts, where a sequence upstream of the start codon is complementary to a downstream sequence (see section 1.8.3). In both cases, the translation initiation region could be sequestered by base-pairing with the downstream antisense-like element, suggesting similar mechanisms for control. Specific proteins might stabilise the secondary structures under non-inducing conditions, or actively resolve them under inducing conditions. The authors suggested various signals that may be involved in the transcriptional and posttranscriptional control of *rpoS* expression. Recently Schweder *et al.* (1996) have reported an increase in stability and levels of σ^S in exponentially-growing *clpXP* cells, suggesting that the protease ClpXP plays a role in the degradation of this sigma under normal growth conditions. These results suggest that the mechanisms underlying the control of σ^S activity under various conditions are very complex.

Within the large σ^S regulon, differential regulation has been observed for subsets of genes, for example in response to anaerobiosis (Atlung and Brønsted, 1994; Brønsted and Atlung, 1994), oxidative stress (Altuvia *et al.*, 1994; Lomovskaya *et al.*, 1994) and osmotic shock (reviewed in section 1.10.2). These observations indicate that additional factors besides σ^S participate in the fine regulation of these genes.

H-NS and σ^S

The histone-like protein H-NS has been shown to be another component of the complex network regulating σ^S and many σ^S-dependent genes. H-NS is a small, abundant DNA-binding protein with a preference for AT-rich curved DNA and is implicated in chromosome organisation due to its ability to negatively supercoil and compact DNA (Lucht *et al.,* 1994; Owen-Hughes *et al.*, 1992; Tupper *et al.,* 1994; Zuber *et al.*, 1994). In addition, it represses the expression of numerous genes with seemingly unrelated functions (Ueguchi and Mizuno,1993; Yoshida *et al.*, 1993). In the case of the *proU* operon, binding of H-NS to the promoter region has been demonstrated (see section 1.10.3).

Yamashino *et al.* (1995) observed enhanced accumulation of σ^S during the exponential growth phase in cells carrying a *hns* deletion, while Barth *et al.* (1995) demonstrated increased expression of at least 22 σ^S-dependent proteins in a *hns* mutant. These authors independently found an increase in σ^S concentration resulting from changes in both translational efficiency and protein stability with little or no effect of H-NS on *rpoS* transcription. These findings implicate H-NS deeply in the posttranscriptional regulation of *rpoS*.

The precise role of H-NS in *rpoS* expression is currently under investigation. It may directly negatively regulate transcription of genes involved in efficient *rpoS* mRNA translation, or of genes involved in σ^S stability such as enhancers of resistance to ClpXP activity. It might even bind directly and specifically to the *rpoS* transcript, since H-NS is known to bind to RNA (Lammi *et al.,* 1984) and can inhibit *in vitro* translation (Spurio *et al.,* 1992). The mode of H-NS derepression that allows the accumulation of σ^S in response to different stimuli remains unclear.

The role of σ^S-independent H-NS repression of osmotically-inducible genes is discussed in section 1.10.3.

σ^S-dependent promoters

In contrast to other minor sigma factors, $E\sigma^S$ is able to recognise many $E\sigma^{70}$-dependent promoters (Tanaka *et al.,* 1993). The promoter specificities are clearly different, since each will activate transcription from certain promoters not recognised by the other. However, alignment of σ^S-dependent promoters has failed to yield a consensus sequence. Most appear to have an $E\sigma^{70}$-type -10 hexamer sequence, but with little homology at -35. Espinosa-Urgel and Tormo (1993) suggested a role for DNA curvature in $E\sigma^S$/promoter recognition, which perhaps aids binding by $E\sigma^S$ when no -35 element is present. However, there is no direct evidence for this, and the bends may be cis-regulatory elements which allow H-NS to bind and repress transcription, as is the case with certain osmotically-inducible genes (section 1.10.3).

The $E\sigma^S$-dependent *fic* promoter does not contain any distinct -35 element, and is recognised weakly by $E\sigma^{70}$. Using hybrid -35/-10 promoters, Tanaka *et al.* (1995) showed that the -10 and downstream region of the *fic* promoter functioned as an efficient promoter for $E\sigma^S$ *in vitro.* This suggests that the discrimination signal of promoters recognised by $E\sigma^S$ alone resides in this region alone, with no apparent role for any upstream bends. Although *fic* transcription is exclusively dependent on *rpoS* *in vivo* (Utsumi *et al.,* 1993), $E\sigma^{70}$ can recognise its promoter *in vitro* at high RNAP concentrations (Kusano *et al.,* 1996), indicating the involvement of additional factors *in vivo*. Indeed, Ding *et al.* (1995) showed that the promoters of the osmoregulated genes *osmB* and *osmY,* which are recognised equally by $E\sigma^S$ and $E\sigma^{70}$ under standard *in vitro* conditions (50mM NaCl), were transcribed at levels 20-fold higher by $E\sigma^S$ at high glutamate concentrations. These conditions mimic the transient increase in potassium glutamate levels upon exposure to high osmolarity (see section 1.10). Thus, a consensus promoter sequence may not exist, if growth-phase or stress-induced factors determine which particular sets of σ^S-dependent promoters are transcribed. Modification of core RNAP is thought to be involved in stationary-specific transcription regulation (Ozaki *et al.,* 1991 and 1992), and variations in core conformation may also direct $E\sigma^S$ to transcribe from other stress-induced promoters. One of the most striking differences between σ^S and the other sigmas (including σ^{70})

is the long tail beyond 4.2 (Lonetto *et al.,* 1992). It is possible that this is involved in interactions with σ^S-specific activator proteins binding to DNA upstream of the relevant promoters, or to σ^S-specific upstream DNA sequences or conformations.

Kusano *et al.,* (1996) investigated the effect of DNA supercoiling on $E\sigma^S$ transcription, since DNA superhelicity is known to be affected by cell growth conditions. For instance, nutrient downshift and stationary growth phase cause a decrease in DNA superhelical density (Balke and Gralla, 1987; Jaworski *et al.,* 1991) while high osmolarity leads to an increase in the superhelicity (Higgins *et al.,* 1988; Hsieh *et al.,* 1991). The *in vitro* results of Kusano *et al.* (1996) indicated that transcription by $E\sigma^S$ was maximal with promoter template DNA prepared at stationary phase, and suggested that $E\sigma^S$ has a higher affinity for DNA with low superhelical density. In addition, they found that transcription from the *osmY* promoter on a highly supercoiled plasmid was maximal at high levels of potassium glutamate. This may allow the σ^S-dependent transcription of certain genes following osmotic upshift when supercoiling increases, while other σ^S-dependent genes remain silent.

The global regulation of transcription by $E\sigma^S$ is complex. Points of control at the levels of transcription, translation and σ^S stability determine the amount of active σ^S within the cell. Additional factors present as a result of various growth and environmental stimuli determine which subsets of genes are expressed. This regulatory strategy presumably allows a degree of specific fine modulation of σ^S-controlled genes with respect to the time of entry into stationary phase and in response to additional environmental conditions.

1.8.2 Sigma F

Helmann and Chamberlin (1988) proposed that a novel sigma might be required for transcription of flagellar and chemotaxis genes in *E.coli*, based on the observation that such operons were preceded by sequences resembling SigD-dependent promoters. SigD regulates the expression of flagellar genes in *B.subtilis*. Arnosti and Chamberlin (1989) isolated RNAP holoenzyme containing a novel 28kDa sigma, required for the transcription of some *E.coli* flagellar and chemotaxis genes, which they named σ^F. σ^F is the same size as SigD and the two sigmas are 37% homologous (Ohnishi *et al.*, 1990). In addition, the cloned SigD gene can restore most *rpoF* regulated functions (except chemotaxis) to a σ^F-deficient *E.coli* strain (reviewed by Helmann, 1991).

σ^F participates in a regulatory cascade of gene expression of more than 40 structural genes required for assembly of flagellae, motility, and chemotaxis (Helmann, 1991). Flagellar and chemotaxis genes are expressed in three temporal stages. The two class I genes encode a pair of transcriptional activator proteins, FhlC and FhlD. Transcription of *fhlC* and *fhlD* is activated by cyclic AMP and the cAMP-receptor protein (CRP). *rpoF* is one of 28 class II genes (which also include the structural genes for the base and hook of the flagellum) dependent on activation by FhlC and FhlD. The α-CTD of RNAP (section 1.2.1) is required for activation, indicating that the FhlC/FhlD complex is a classI transcriptional activator (Liut *et al.*, 1995). σ^F is required for transcription of most of the 18 class III structural genes following the expression of the class II genes. The coordinate regulation of *rpoF* and the base/hook genes ensures that class III gene expression (and in particular the very costly production of flagellin) does not occur wastefully prior to assembly of a complete base-hook structure. A negative regulatory locus, *flgR*, has been identified in *S.typhimurium*, which may block σ^F activity in the absence of complete class II gene expression. Mutations inactivating *flgR* allow class III gene expression in the absence of many class II gene products (Gillen and Hughes, 1991).

Putative σ^F-dependent promoters have been identified upstream of *rpoF* and many other class II genes but these are not essential for their expression *in vivo*. Helmann (1991) suggested that these promoters may facilitate σ^F-dependent amplification of class II gene expression concomitant with expression of class III genes. Interestingly many of these promoters lack good -35 promoter elements, a feature reminiscent of some positively activated σ^{70}-dependent promoters, suggesting that other positive factors may be involved in late class II gene expression.

1.8.3 Sigma 32

When cells are shifted from 30°C to 42°C, synthesis of well-conserved heat-shock proteins (HSP) is markedly and transiently induced. The role of HSP is to assist protein folding, assembly, transport, repair and degradation during normal growth as well as under stress conditions. In *E.coli* the *rpoH* gene product σ^{32} is specifically required for the expression or increased expression of over 30 genes encoding HSP such as the GroEL and DnaK chaperones, and Lon and Clp proteases (Hendrich and Hartl, 1993), while the σ^E regulon activates additional genes necessary for similar functions in respect to extracytoplasmic proteins (section 1.8.4). Ordinarily σ^{32} is very unstable, with a half-life of one minute, and its synthesis is largely repressed. Upon heat-shock the σ^{32} levels are increased transiently by both stabilisation and enhanced synthesis (Straus *et al.,* 1987). In parallel, the rate of HSP synthesis increases 10- to 15-fold within five minutes. The unusual instability and restricted synthesis of σ^{32} provide an effective means for keeping HSP levels to a minimum during normal growth and for rapid adjustment upon exposure to stress.

When downshifted from 42°C to 30°C expression of heat-shock genes decreases about 10-fold within five minutes before reaching a new steady-state level. In this case, the rapid drop in heat-shock gene expression results from a decrease in σ^{32} activity rather than in its amount (Gross *et al.,* 1990). These findings show the need

for tight regulation of the HSP levels to assure maximal growth and survival, which is controlled by modulating the synthesis, degradation, and activity of σ^{32}.

rpoH is served by at least 4 promoters (P1, P3, P4, and P5). Three (P1, P4, and P5) are recognised by $E\sigma^{70}$ whilst P3 is recognised by $E\sigma^{E}$ (section 1.8.4). The major promoters P1 and P4 are responsible for *rpoH* transcription under most conditions, whilst $E\sigma^{E}$ has been shown to initiate at P3 only at very high temperatures (45-50°C) where σ^{70} is largely inactive (Gross *et al.,* 1990). Despite the complex and potential regulatory importance of multiple promoters, the bulk of the transient increase in σ^{32} levels observed upon heat-shock results from increased translation rather than from increased transcription (reviewed by Yura *et al.,* 1993). Transcriptional control of σ^{32} appears to be important only in maintaining basal levels of *rpoH* mRNA under a variety of steady-state conditions and under extreme conditions at 45-50°C.

Extensive analysis of deletion derivatives of an *rpoH::lacZ* gene fusion led to the identification of two 5′-proximal regions, designated A and B, of *rpoH* mRNA involved in thermoregulation (Nagai *et al.,* 1991). Region A is a positive element located near the initiation codon which enhances translation, probably through its complementarity to a 16S rRNA sequence, somewhat upstream of the Shine-Dalgarno complement at the 3′ end. Region B is a long internal negative element which represses translation during normal growth. Random point mutations which cause high constitutive expression of σ^{32} were found to occur specifically within these two regions (Yuzawa *et al.,* 1993). Nagai *et al.* (1991) predicted that a fairly stable mRNA secondary structure is formed, spanning the initiation codon plus region A and part of region B. The secondary structure as well as specific base pairing appears to be important for temperature regulation, suggesting that a *trans*-acting binding factor(s) modulates the rate of translation (Yuzawa *et al.,* 1993).

Translational shutoff of σ^{32} requires additional *cis*- and *trans*-acting factors. DnaK, DnaJ, or GrpE chaperones exert a negative feedback control on the synthesis of σ^{32} and HSP since mutations in *dnaJ, dnaK,* and *grpE* resulted in enhanced heat-shock gene expression at low temperatures and an extended heat-shock effect following temperature upshift (Tilly *et al.,*1983; Straus *et al.,* 1990). Moreover, deletion

analysis of an *rpoH::lacZ* gene fusion showed that a specific internal segment of σ^{32} called region C was required for normal repression (Nagai *et al.,* 1994). The amino acid rather than nucleotide sequence was found to be important, indicating the involvement of protein-protein interaction in translational repression. Since DnaK/DnaJ chaperones are known to interact with σ^{32} *in vivo* and *in vitro* (Gamer *et al.,* 1992; Liberek *et al.,* 1992), region C might provide sites for such interaction resulting in translational shutoff. Consistent with this idea, two motifs with high affinity for DnaK (but not DnaJ) have been identified in region C (McCarty *et al.,* 1996). These motifs were found to be highly conserved among σ^{32} homologues but missing in other sigmas, indicating that region C is central to a regulatory mechanism allowing the sensing of stress by the heat-shock transcription complex. It appears that the DnaK/DnaJ/GrpE chaperones contribute to the characteristic instability of σ^{32} by facilitating proteolysis through an interaction with σ^{32} or a protease(s) responsible for its degradation. Tomoyasu *et al.* (1995) demonstrated that a temperature-sensitive *ftsH* mutation stabilised σ^{32} at high temperatures, and purified FtsH protein was shown to degrade σ^{32} in an ATP-dependent manner, suggesting that this membrane-bound metalloprotease plays a major role in the turnover of σ^{32} and in the heat-shock response.

Inactivation of σ^{32} contributes to the rapid shutoff of HSP synthesis early during the adaptation phase (Yura *et al.,* 1993). A DnaK/DnaJ complex is thought to be involved in sequestering or inactivating σ^{32}, since DnaK and DnaJ compete with core RNA polymerase for binding to σ^{32} (Gamer *et al.,* 1996). Heat-shock is thought to increase cellular levels of partially unfolded or misfolded proteins that should preferentially bind to DnaK and/or DnaJ. This would reduce the levels of free chaperones, which should in turn stabilise σ^{32}, thus leading to HSP induction. When enough free chaperones accumulate, they would then exert a negative feedback control on the synthesis, stability, and activity of σ^{32}. Further dissection of the regulatory circuits of σ^{32} should provide useful information regarding the exact mechanisms that a cell employs to sense and respond to an increase in temperature.

1.8.4 Sigma E

Transcription from the *rpoH* P3 promoter, which is active at extreme high temperature in *E.coli*, was found to be independent of both $E\sigma^{70}$ and $E\sigma^{32}$ (Erickson *et al.,* 1987). The factor responsible for *rpoH* transcription at high temperature (50°C) was purified by Wang and Kaguni (1989) and Erickson and Gross (1989). Both groups identified it as a novel form of RNAP containing a sigma of 24kDa. The protein was named σ^E to denote that it is active at extreme high temperature. Another gene known to be transcribed by $E\sigma^E$ is *degP* (*htrA*) (Erickson *et al.,* 1987) which encodes a periplasmic protease essential at temperatures above 42°C (Lipinska *et al.,* 1988,1989). The periplasmic location of DegP suggests that the $E\sigma^E$ regulon may protect extracytoplasmic compartments from stress. Indeed, Mecsas *et al.* (1993) found $E\sigma^E$ activity increased following overproduction of outer membrane proteins, and decreased when these proteins were underproduced. Mutations resulting in misfolded outer membrane proteins also induced $E\sigma^E$ activity whilst having no effect on $E\sigma^{32}$. σ^E belongs to the Extracellular Function (ECF) subgroup of sigma factors, which is quite divergent from other sigmas of the σ^{70}-type family (Lonetto *et al.,* 1994). All the sigmas in this subgroup appear to regulate extracytoplasmic functions and respond to extracytoplasmic signals. Mecsas *et al.* (1993) proposed that the two heat-inducible sigmas may have distinct but complementary roles in monitoring the state of proteins in the cytoplasm (σ^{32}) and outer membrane and periplasm (σ^E).

Little is known about the σ^E regulon and its control. Analysis of transcription of the cloned *rpoE* gene identified two major transcripts (Rouvière *et al.,* 1995; Raina *et al.,* 1995). Although the upstream startpoint was not located near any obvious canonical σ^{70}, σ^{32}, or σ^E-dependent promoters, the second startpoint was located just downstream of a sequence with high similarity to the two previously identified σ^E promoters. Both groups independently confirmed that $E\sigma^E$ can transcribe *rpoE* both *in vivo* and *in vitro*, and that constitutive transcription at 50°C originates primarily at the downstream promoter. In addition to this autoregulation it seems likely that other

elements are involved in controlling *rpoE* expression. Most of the ECF sigmas examined are in operons that also encode negative regulators of the sigma protein's activity. In particular AlgU, the ECF sigma most closely related to σ^E, is negatively regulated by two proteins called MucA and MucB, also encoded by the *algU* operon (Deretic *et al.,* 1994). Rouvière *et al.* (1995) noticed that the DNA sequence immediately downstream of *rpoE* encodes a protein with ~40% identity to MucA, suggesting that σ^E may also be negatively regulated by an antisigma encoded in its own operon.

Overproduction of σ^E led to the increased synthesis of at least 11 proteins including σ^E, σ^{32}, and DegP (Rouvière *et al.,* 1995; Raina *et al.,* 1995). At 50°C almost all *rpoH* synthesis is carried out by $E\sigma^E$ (Erickson and Gross, 1989), and so it affects the production of HSP by the σ^{32} regulon. However, it seems likely that the σ^E regulon provides additional functions necessary for growth under more extreme conditions. Characterisation of the other members of this regulon should provide further insight into the global regulation of the heat-shock response.

1.8.5 FecI

Although the sequence of *fecI* was published in 1990 (van Hove *et al.,* 1990), FecI was not then identified as a potential sigma factor. However, its homology with members of the ECF subgroup of the σ^{70}-type family later indicated that it is a sigma (Lonetto *et al.,* 1994). FecI was originally characterised as being essential for the induction of the ferric citrate transport genes encoded by the *fec* operon in *E.coli.* Briefly, ferric citrate binds to the outer membrane protein FecA, and crosses the membrane via TonB, ExbB, and ExbD (Zimmermann *et al.,* 1984). The C-terminal portion of the cytoplasmic membrane-bound protein FecR responds to the ferric citrate, and the signal is transduced to FecI, located in the cytosol, by the N-terminal portion of FecR (Ochs *et al.,* 1995). FecI activates transcription of the *fecABCDE* transport genes without the need for ferric citrate to enter the cell. Uptake of ferric

citrate through the cytoplasmic membrane accordingly ensues, following the synthesis of the transport proteins FecB/FecC/FecD/FecE (Schultz-Hauser *et al.*, 1992).

The *fec* genes are arranged *fecIRfecABCDE* and are transcribed in this order from a σ^{70}-dependent promoter upstream of *fecI* and the FecI-dependent promoter upstream of *fecA* (Enz *et al.*, 1995). FecI is only weakly active in the absence of ferric citrate and FecR, although overproduction leads to low constitutive *fec* expression (Ochs *et al.*, 1995). Within the *fec* operon the levels of *fecBCDE* RNAs are low compared to the level of *fecA* mRNA, and processing from the 3′ end is proposed to stop near the end of *fecA* where a hairpin structure has been identified (Enz *et al.*, 1995).

Transcription of the *fec* operon is repressed by the Fur repressor loaded with Fe^{2+} (Zimmermann *et al.*, 1984) and nucleotide sequences typical for Fe^{2+}-Fur binding sites have been located in the *fecI* and *fecA* promoter regions (Enz *et al.*, 1995). In addition, the FecI-dependent promoter of *fecA* shown in table 1.3 was found to have sequence similarity to promoters of genes regulated by other ECF sigma factors, including the σ^{E}-dependent *rpoH* P3 promoter (Enz *et al.*, 1995). FecI was subsequently shown to function as a sigma since purified protein in conjunction with RNA polymerase core enzyme specifically initiated transcription from the *fecA* promoter *in vitro* (Angerer *et al.*, 1995). No specific transcript was obtained with RNAP carrying σ^{70}, and σ^{70} competed with FecI for core binding. Furthermore the electrophoretic mobility of DNA fragments carrying the upstream region of *fecA* was retarded by RNA polymerase core enzyme after addition of FecI. Ochs *et al.* (1995,1996) demonstrated that point mutations in the FecI HTH motif either increased or decreased the induction of transcription of a *fecAB*::*lacZ* fusion, and that chromosomal insertion or deletion mutants in *fecI* were totally devoid of FecA production and *fecAB*::*lacZ* expression. Ochs *et al.* (1996) also demonstrated that *fecI* is not autoregulated, but is regulated by iron via Fe^{2+}-Fur.

FecI is unique amongst *E.coli* sigma factors in that it only directs transcription from one known promoter. It may be that other proteins originally identified as activators of transcription can actually function as sigma factors.

1.8.6 Sigma 54

σ^{54} stands alone among the alternative sigma factors in exhibiting no significant amino acid homology to the σ^{70} family (section 1.7). Wong *et al.* (1994) made a series of deletions in *rpoN* and analysed the properties of the mutant proteins. Their results identified a hydrophobic heptad repeat (HHR) sequence (fig. 1.2) important for binding to core RNA polymerase. PCR mutagenesis confirmed the requirement for residues in this region, which are clustered in a short sequence speculated to be homologous to a segment of the core binding region of σ^{70} (Tintut and Gralla, 1995; Lesley and Burgess, 1989). Further determinants of σ^{54} binding were also found in flanking regions which have no homology with σ^{70}, suggesting that each sigma may bind multiple domains on core, only one of which is common in this case.

The deletion studies of Wong *et al.* (1994) showed that a large carboxyl section of the protein, including a putative HTH motif (fig. 1.2), was essential for -24 promoter element recognition and DNA binding. A smaller amino-terminal segment was found to be essential for DNA melting and -12 promoter element recognition, but playing a secondary role with respect to DNA binding. The amino terminus contains acidic, glutamine-rich and periodic leucine-rich motifs which overlap, making the identification of potential roles for these motifs difficult. However, point mutational analysis has implicated the glutamine-rich motif in assisting DNA melting, whilst leucine residues comprising a heptad repeat structure have been shown to be critical for -12 promoter element recognition and mRNA production (Hsieh and Gralla, 1994).

As mentioned in section 1.7, members of the σ^{54} family can bind to promoter DNA in the absence of core enzyme. Tintut *et al.* (1995) showed that σ^{54} remained

at the promoter whilst core polymerase catalysed RNA elongation, probably due to tight contacts between the sigma and promoter DNA. The authors proposed a novel transcription cycle distinct from that mediated by σ^{70}. In this model, the closed complex containing the holoenzyme forms first and is converted to an open complex by the binding of an activator protein. In the presence of ribonucleotides, polymerase initiates transcription and proceeds downstream whilst σ^{54} is left at the promoter. There then appears to be a requirement for σ^{54} to be released or to undergo a change in conformation prior to reinitiation. This allows another possible step at which transcription might be controlled by allowing different mechanisms for initiation and reinitiation. Induction may be controlled via the activator protein, whilst reinitiation may be controlled separately by factors that work on the bound sigma. This alternative transcription cycle may reflect the different properties of σ^{54} and σ^{70}. The σ^{70} domain structure ensures that most promoters will not be bound by polymerase prior to their required activation. Because so many promoters are σ^{70}-dependent, this property allows for optimal use of a limited amount of RNAP. In contrast, the less common σ^{54}-dependent promoters are possibly permanently bound by this sigma, which might be more appropriate to its unusual function in permitting activator-dependent transcription of a relatively small number of genes.

Detailed structure/function analysis of σ^{54} is still at an early stage, but further mutational analysis should reveal which regions are important for specific functions of this unique sigma factor.

1.9 Structure and Function of Bacterial Interspersed Mosaic Elements

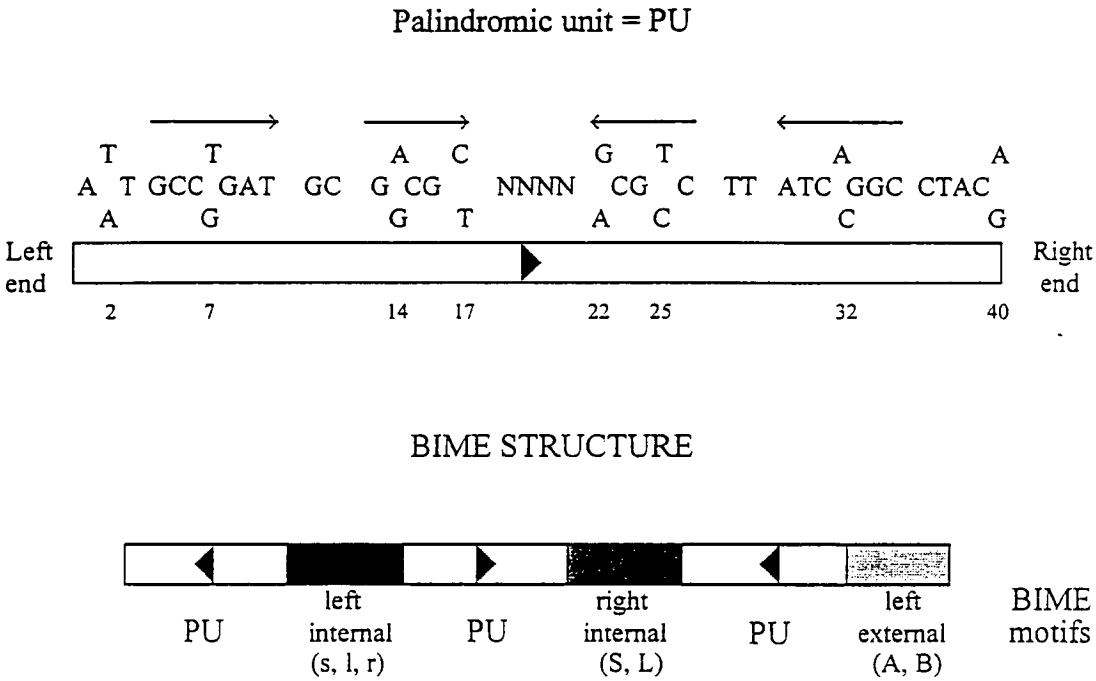
The presence of different families of interspersed DNA sequences located outside structural genes appears to be a general phenomenon in bacterial genomes (Blaisdell *et al.*,1993). They include Bacterial Interspersed Mosaic Elements, or BIMES (Gilson *et al.*, 1991), which are 40 to 500-basepair long DNA sequences found in several enterobacterial species (Bachellier *et al.*, 1993). They are composed of several motifs, including a 40-basepair sequence with imperfect dyad symmetry called PU for Palindromic unit (Gilson *et al.*, 1984; 1987; Higgins *et al.*, 1982) or REP for Repetitive Extragenic Palindrome (Stern *et al.*, 1984). Although not found in protein-coding regions, they are usually located in transcribed regions of the chromosome.

1.9.1 Palindromic Units

PU elements are dispersed around the chromosome, and make up almost 1% of *E.coli* genomic DNA (Gilson *et al.*, 1984; 1987). Analysis of different PU sequences yielded a consensus (fig. 1.3). The highly conserved dyad symmetry could allow the formation of a strong stem-loop in the transcribed form, and evidence for such RNA secondary structure has been obtained (Higgins *et al.*, 1988).

PU elements often appear in clusters of two to six but with one occurrence of ten and one of eleven known to date (Bachellier *et al.* , 1994). Because the palindrome is imperfect the PU can be oriented. Strikingly, successive PU within a cluster invariably alternate in orientation. In addition, where a base in the left arm of the dyad symmetry differs from the consensus, the corresponding base in the right arm also differs so as to maintain base-pairing potential.

Figure 1.3. Structure of a BIME. The PU consensus is as previously determined (Gilson *et al.*, 1987). The arrows indicate the positions that exhibit dyad symmetry. The PU is symbolised below the consensus by a white rectangle with a triangle to indicate the orientation, from left to right, according to the positions not included in the symmetry (numbers 11-12 and 27-28). Variable positions are numbered below the diagram. The same PU symbol is used, along with the extra-PU motifs, to illustrate the structure of a BIME. The external motifs flank the left end of a single PU while the internal motifs are located between two successive PU occurences. The internal motifs have been further subdivided according to their sequence and to the ends of the adjacent PUs. See text for explanations of symbols s, l, r, S, L, A, B.



1.9.2 Extra-Palindromic Motifs

Detailed analysis of DNA sequences surrounding PU elements revealed seven other conserved motifs (Gilson *et al.*, 1991). PU clusters were divided into segments according to their positions relative to the PU, and the lengths and sequences of segments occupying similar positions were compared. Segments located at each end of a PU cluster were called external segments. Because of the orientation of PU elements, right external sequences which flank the right end of a PU were distinguished from left external sequences, which flank the left end of a PU (fig.1.3). Segments located between two PU were called internal segments. Because of the strict alternation in PU orientation within a cluster, right internal segments which are located between the right ends of two adjacent PU were distinguished from left internal segments, located between the left ends of adjacent PU (fig. 1.3).

Two sets of similar sequences for left external segments were found, with homogeneity averaging 55% for group A segments and 60% for group B segments. By contrast, no such sets were found for right external segments. Sorting of internal segments into sets based on size and sequence identified five motifs. About 50% of right internal segments were found to have a size ranging from 32 to 34 bp, and were called L for Long sequences. The L set exhibits high sequence homogeneity, averaging 80%. The remaining segments, between 11 and 14 bp in length, were called S for Short sequences. Sequence homogeneity averages 50%. Size distribution revealed three sets of left internal segments. l for long sequences are 8 or 9 bp in length and average 68% sequence identity. s for short sequences are composed of a single basepair, which is C:G in 60% of the cases. The third set, r, range in size from 18 to 31 bp, with no sequence identity. It appears that PU clusters are composed exclusively of a mosaic combination of the PU sequence itself with these seven surrounding motifs. Within any single PU cluster, all the right internal segments are either S or L, while all the left internal segments are either s, l, or r. Combination of the eight motifs can generate a great diversity of structures, which became known as BIMEs, for Bacterial Interspersed Mosaic Elements.

1.9.3 BIME Families

More recently it has been shown that the *E.coli* PU motif can be subdivided into three classes, Y, Z^1, and Z^2 (Bachellier *et al.*, 1994). While the earlier PU consensus allowed for considerable variation, the newly defined PU motifs are highly homogeneous. PU sequences were grouped according to the presence of GC or TA at variable positions 7 and 32 (fig. 1.3), and called respectively the Y or Z type. The majority of Y-type sequences are 38 bp long with a high sequence homogeneity averaging 92%. The majority of Z^1-type sequences are 33 bp while the majority of Z^2-type sequences are 40 bp in length. Both Z motifs also exhibit a high level of sequence homogeneity, averaging 94% for Z^1 and 93% for Z^2. The sequences are shown in figure 1.4. Within each class of PU motif the loop region, which was previously thought to be highly variable, appears to be as homogenous as the rest of the PU sequence. Available genomic DNA sequences indicate that the three PU motifs are equally distributed on the bacterial chromosome.

A search for preferential association between the newly recognised PU motifs and the seven extra-palindromic motifs showed that when association with a right internal sequence is examined, 81% of the Z^1 motifs are associated with the L motif, while 80% of Z^2 motifs are associated with the S motif. Preferential association is also found with the left internal sequences : the Z^1 motifs are associated with l and r motifs, while the Z^2 motifs are mostly associated with l and s motifs. In addition, it was also found that the left external motif A is always associated with the Z-type PU, and that the left external motif B is always associated with the Y-type PU. Furthermore, 89% of Z^1 motifs are in BIMEs containing only two PU; among them, 49% have the structure AZ^1LYB, refered to as BIME-1. In contrast, 50% of the Z^2 motifs are in BIMEs with more than 2 PU; among them, 65% have the structure $(YSZ^2(s,l))_n$ with n equal to or greater than 1.5, and are refered to as BIME-2. These results show that the association between motifs within BIMEs is not random, and that two major families of BIME are present in *E.coli*.

Figure 1.4. Sequence alignment of the three PUs. Regions of similarity have been numbered from I to V by Bachellier *et al.* (1994) and are underlined below. The asterisks indicate positions that are different between at least two of the three PU motifs; dashes represent gaps introduced to obtain an optimal alignment.
The consensus sequence of the transcriptional terminator PU* (Gilson *et al.*, 1986 and 1987) is also given.

```
      ***   I   *    II      ****  ****  III *    IV   *      V

      AA                            G                          A
Y       T GCC G GATGCG GCG-T   AA- CG C CTTATC C GGCCTAC
      CT                            A                          G

                                                               A
Z1    ATT GCC T GATGCG - - - CTA - - - CG - CTTATC A GGCCTAC
                                                               G

      AAT                           G                          A
Z2      GCC T GATGCG ACGCT  GCGCG T CTTATC A GGCCTAC
      GTC                           T                          G

PU*       GCC G GATG       GGATATCCA      C ATC C GGC
```

1.9.4 Generation of BIMEs

The structures of other BIMEs, corresponding mostly to ones containing a single PU, can be derived from the structures of the main families. Thus it is likely that BIME-1 and BIME-2 correspond to two ancestral BIME copies. Different mechanisms have been proposed for BIME formation: transposition (Gilson *et al.*, 1984), retrotransposition (Gilson *et al.*, 1990 ; Higgins *et al.*, 1988 ; Richetti and Buc, 1993), or formation of stem-loop DNA fragments (Ohshima *et al.*, 1992). Whether the two types of BIME sequences are formed by the same mechanism is not known. Since the same PU and internal motifs are strongly conserved within a particular BIME, it seems that they originate from an array of direct repetitions. This could be explained by a localized amplification of a PU doublet. Interestingly, DNA polymerase I, which is known to bind BIME sequences (section 1.9.7iii), is also believed to be involved in some amplification processes (Kornberg, 1980). Perhaps specific BIME-DNA polymerase I interaction can lead to localised amplification.

1.9.5 Functions of BIMEs

It has been estimated that the *E.coli* chromosome carries 300-500 BIMEs, but at present their biological function(s) is unknown. Roles have been proposed in the regulation of gene expression, in chromosome architecture, and in chromosomal rearrangements. These are discussed in detail in the following sections.

1.9.5a Regulation of gene expression

mRNA stabilization

Newbury *et al.* (1987) found that in the *malEFG* operon of *E.coli,* upstream *malE* mRNA is considerably more stable than full-length *malEFG* mRNA. The *malEFG* operon is one of two divergent operons that comprise the *malB* locus encoding the

maltose/maltodextrin transport system (Hengge and Boos, 1983). The *malE* gene encodes an abundant periplasmic protein, present in a 20- to 40-fold excess over the membrane associated MalF and MalG proteins (Manson *et al.*, 1985). It was also noted that a BIME-2 sequence including two PU elements is located in the *malE-malF* intergenic region. Upstream mRNA can be stabilized by impeding the action of 3′-5′ exonucleases, which is frequently accomplished by stable secondary (stem-loop) structures. It was therefore suggested that the BIME was responsible for the increased stability of upstream *malE* mRNA and hence, in part at least, for the differences in efficiency of gene expression within this operon. To test this hypothesis Newbury *et al.* (1987) deleted one or both of the PU elements on the chromosome. They found that with both PU elements deleted, the stability of the upstream transcripts was reduced such that *malE*-specific transcripts no longer accumulated. However, mRNA stability was unaffected when only one PU was removed. Correspondingly, whereas removal of both PU elements reduced the rate of MalE protein synthesis by nine-fold, removal of only one had no effect.

Since the ratio of MalE to MalF proteins has been estimated at 20-40:1, the difference between upstream and downstream expression cannot be due to BIME-dependent mRNA stabilization alone. Other factors may include differences in ribosome binding and translational efficiency between *malE* and *malF* mRNA. However, these results show that BIME-mediated differential mRNA stability can play a significant role in determining the relative expression of genes within an operon.

Control of Translation

Stern *et al.* (1988) went on to suggest that in operons with unequal gene expression, intercistronic BIMEs may interfere with efficient downstream translation as well as stabilise upstream mRNA. Working with the histidine permease operon of *Salmonella typhimurium*, which has an intercistronic region containing two PU elements located between the first gene (*hisJ*) and three distal genes (*hisQMP*), they

found that deletion mutants lacking the PU elements but retaining the *hisQ* Shine-Dalgarno sequence increased the level of *hisQ* expression almost four-fold relative to wild-type. This supported a previous observation by Stern *et al.* (1984) who reported a decrease in expression of *galK* on a plasmid vector when the same PU sequences were inserted upstream of the reporter gene. The mechanisms involved remain unknown.

Transcriptional Termination

Despite their locations exclusively outside structural genes, their ability to form stem-loop structures, and their association with several proteins (next section), BIMEs do not appear to play any role in transcription termination. This is consistent with their presence in intercistronic regions. However, Gilson *et al.* (1986 and 1987) showed that a subclass of PU called PU* (fig. 1.4) was able to act as a bidirectional terminator of transcription from the convergent genes *pheA* and *tyrR*. Surprisingly, the *pheA* terminator has no 3′ run of uridines as seen in factor-independent transcriptional terminators (the *tyrR* terminator has 4). The 6 known PU* sequences are all located between two convergent genes and make up most of the DNA in these regions. No other potential transcriptional terminators were found in their vicinity, leaving open the possibility that they all have this function. As yet, no further information is available regarding the frequency of PU* on the chromosome.

Although the above results indicate that BIME structures play a part in the regulation of gene expression, some of the effects can be explained as consequences of the palindromic secondary structures forming in mRNA. Moreover, extensive sequence homology is not usually found among RNA structures involved in a generalized function (e.g. Rho-independent transcriptional termination or RNase protection structures), whereas the PU sequences retain remarkable sequence homology. Therefore, the above may not be the primary reasons for the maintenance of BIMEs on the chromosome.

1.9.5b Chromosome Architecture

The palindromic nature and conservation of BIMEs suggests that they could be specifically recognized by proteins. Indeed, a role in chromosome structure and organization has been suggested, based on the nature of several proteins which associate with BIMEs. The chromosome of *E.coli* K12 is a single negatively supercoiled covalently closed circle comprising about 4.72Mb. Exponentially growing cells contain two to four chromosomes which are condensed into nucleoid structures that are capable of carrying out rapid replication and chromosome segregation. Biochemical studies and electron micrographs suggest that the nucleoid is organized into large independent loops whose structure is influenced by the action of topoisomerases and by the binding of histone-like proteins (reviewed by Drlica, 1990). These individually supercoiled domains are thought to be anchored either to each other or to the membrane by way of specific interactions at the bases of the loops.

DNA gyrase and HU binding

Yang and Ames (1988) showed that the PU motif is bound specifically by DNA gyrase *in vitro,* and that gyrase binds progressivelly better to DNA containing one, two, and four PU sequences. In addition, they found that the histone-like protein HU stimulated the binding of gyrase to PU, and also reduced the level of gyrase-mediated DNA cleavage.

HU constitutes a major fraction of DNA-binding proteins in isolated bacterial nucleoids and has been found to affect site-specific interactions of other proteins involved in DNA transposition (Craigie *et al.,* 1985), site-specific recombination (Johnson *et al.,* 1986), and DNA replication (Dixon and Kornberg, 1984; Ogawa *et al.,* 1985). HU preferentially binds to supercoiled DNA rather than to relaxed DNA (Gualerzi *et al.,* 1986), and it has been shown to wrap the DNA into a nucleosome-like structure in which negative supercoils are constrained (Broyles and Pettijohn, 1986). HU may play a regulatory role in the interaction of gyrase and PU sequences. Poniggia *et al.* (1993) found that HU prevents the extrusion of cruciforms likely to

be formed by the inverted repeat sequences found in PU elements. By binding to an intermediate in cruciform generation, HU may interfere with the formation of these structures.

Upon identifying the three different PU motifs Y, Z^1, and Z^2 (section 1.9.3), Bachellier *et al.* (1994) studied their relative affinities for gyrase. Using an electrophoretic mobility shift assay, they found the gyrase concentration that led to half binding of a DNA probe carrying a particular PU was lower for Y than for Z^2, (50nM for Y versus 80nM for Z^2), while the concentration for Z^1 was significantly higher (155nM). Since the main differences between Z^1 and Z^2 are in the centre of the PU (fig. 1.4) this region may be critical for gyrase binding. Indeed, the authors found that a centrally deleted form of Y, Y^{Δ} was bound by gyrase with an affinity identical to that of Z^1. It is worth noting that the transcriptional terminator PU* is small (21bp), lacks the central region, and binds gyrase poorly (Gilson *et al.*, 1986; 1987).

Since the central parts of Y and Z^2 share limited sequence homology but are similar in size (fig1.4) it may be that the external parts of the PU, which are highly conserved between the different motifs, are directly recognized by gyrase, but only if the spacing is optimal.

Integration Host Factor binding

Integration host factor (IHF) is a sequence-specific DNA binding and bending protein that is abundant in the *E.coli* cell. The protein is a heterodimer, and belongs to the family of histone-like proteins that share amino acid sequence homology with HU (Drlica and Rouvière-Yaniv, 1987). IHF participates in a number of cellular processes such as λ site-specific recombination, DNA transposition and inversion, phage DNA packaging, plasmid and phage replication and the control of gene expression (reviewed in Freundlich *et al.*, 1992).

IHF binds to specific DNA sequences and protects 30-40 bp against DNaseI digestion. A number of sequences with consensus binding sites for IHF have been suggested. Craig and Nash (1984) found that IHF-binding sites contain a non-

symmetrical consensus 5'- PyAANNNPuTTGATW-3', while Kur *et al.* (1989) proposed a related 27 bp consensus. Computer analysis of *E.coli* sequences to detect potential binding sites for IHF revealed that the highly conserved BIME motif L carries an IHF recognition sequence (*ihf*). BIMEs carrying the L motif flanked by two PU motifs were called "reiterative *ihf* BIMEs" (RIBs) by Boccard and Prentki (1993), or "repetitive IHF-binding palindromic elements" (RIPs) by Oppenheim *et al.* (1993), and belong to the BIME-1 (section1.9.3) family as described by Bachellier *et al.* (1994).Using gel retardation assays and DNaseI footprinting, the above groups independently showed that IHF does bind to this form of BIME *in vitro*. The DNaseI footprint produced by IHF coincided almost exactly with the L motif, with little protection of the surrounding PU sequences. Boccard and Prentki suggested that by binding RIB elements, IHF forms a loop which enhances the interaction of gyrase with the flanking PU sequences.

RIBs/RIPs appear to be evenly distributed around the chromosome and there are estimated to be 50-100 per genome. Interestingly, this coincides with the estimated number of independent domains in the nucleoid (Pettijohn,1982). Thus, IHF-mediated DNA bending at RIBs/RIPs may facilitate the formation of a higher order DNA nucleoprotein structure involved in the organization of the nucleoid, while associated gyrase binding might ensure supercoiling in each domain.

Another possible function of RIBs/RIPs has been suggested in view of an observed bias in their local position: in contrast to the general population of BIMEs, which can be found at intra-operonic locations, RIBs/RIPs are only found at the very end of transcription units. Liu and Wang (1987) have shown that movement of RNA polymerase along a DNA molecule can result in the formation of negative supercoils behind the transcription complex, and of positive supercoils ahead, if the ends of the DNA molecule or the transcription apparatus itself are not free to rotate. RIB/RIP elements are thus ideally located to bind DNA gyrase and function as swivels for the removal of transcription-generated positive supercoils. In support of this, Boccard and Prentkl noticed that at least half of their RIBs were located at the end of transcription units encoding a membrane-associated protein. In such cases, where the

nascent RNA encodes a protein carrying a transmembrane domain which is cotranslationally inserted into the membrane, there could be particularly severe restriction of the rotation of the transcription machinery.

Using an L motif probe against genomic DNA from various enterobacteria, Gilson *et al.* (1991) concluded that this motif, and thus RIBs/RIPs, are present only in *E.coli* and *Shigella sonnei.* This could suggest that these elements are inessential, or that other bacterial species have different elements to carry out the same function. The role played by IHF could be performed by another DNA-bending protein or by an intrinsic bend in the DNA itself. Inspection of *Salmonella* PU-flanking regions did not provide evidence for these hypotheses, but much of the *Salmonella* genome remains unsequenced. Subsequently Bachellier *et al.* (1993) discovered a BIME element in *Klebsiella aerogenes* which fulfils the criteria for RIBs/RIPs, i.e. two convergent PU sequences flanking an L motif. The *Klebsiella* PU and L motif sequences are significantly divergent from those in *E.coli*, but an *ihf* site is present.

The high degree of conservation of the L motif sequence outwith the IHF-binding consensus may suggest possible interaction with other proteins (Herman and Schneider, 1992). An alternative explanation is that RIBs/RIPs have only recently been duplicated and dispersed throughout the *E.coli* genome; hence there has been little time for significant DNA sequence divergence.

DNA Polymerase I binding

Gilson *et al.* (1990) identified two protein fractions purified from an *E.coli* extract which were able to specifically protect PU sequences against ExonucleaseIII (ExoIII) digestion. One of these fractions was shown to be DNA polymerase I (Pol I) judjed by the amino acid sequence of tryptic fragments. Commercially available Pol I gave the same ExoIII protection pattern, although the large C-terminal fragment of Pol I (Klenow protein) did not. Pol I/PU interaction was further confirmed using gel retardation and DNaseI footprint analyses. It was suggested that the PU could serve as preferred entry sites for PolI, playing a role in secondary replication initiation sites, in specific pausing sites during the polymerisation reaction, or in replication

fidelity. The second fraction conferring ExoIII protection contained three major polypeptides of 130kD, 51kD, and 50kD, and two minor polypeptides of 80kD and 66kD, as shown by SDS-PAGE. None of these were identified, but did not comigrate with purified DNA gyrase.

The physiological functions of supercoiling are presumably compaction of DNA and modulation of gene expression. Some promoters are activated by an increase in negative supercoiling, and others are activated by a decrease. In a thermodynamic sense the negative superhelical torsion within the double-stranded DNA helix should favour the equilibrium of a process involved in the unwinding of DNA. It is therefore puzzling that an energetically favourable contribution in the form of superhelical torsion within the DNA helix does not in all cases increase promoter strength. In some cases DNA supercoiling might alter the promoter conformation such that binding by RNAP and/or activator proteins is weakened, resulting in diminished promoter strength. It is not known whether different operons are tuned to respond precisely to the particular degree of supercoiling of the domain in which they are located. It is conceivable that different degrees of superhelicity exist in various chromosomal domains, and that this may be due to the inherent differences between BIME sequences and therefore to variations in their interactions with gyrase, HU, IHF, and/or other BIME-associated proteins.

1.9.5c Chromosome Rearrangements

Chromosome rearrangements giving rise to gene amplification are important evolutionary events (Ohno, 1970) that increase gene expression by increasing gene dosage and/or placing genes under the control of new promoters. They can also create a source of redundant DNA potentially utilisable for generating genetic diversity. The ability to form stable base pairing between homologous regions is thought to be an important feature of the mechanism that gives rise to tandem duplications.

BIMEs could play a role in such rearrangements, and their intergenic location is compatible with this hypothesis. Indeed, several genetic duplications have one end in the PU-containing intercistronic region of the histidine biosynthetic operon of *Salmonella typhimurium* (Anderson and Roth, 1978). Shymala *et al.* (1990) deduced that these duplications arose by recombination with distant PU elements, and showed that recombination between non-identical PUs could take place in a $RecA^-$ background. When a system not involving RecA is responsible for these events, DNA gyrase may play a role, since it can bind PU sequences, and can cause illegitimate recombination (Naito *et al.*, 1984). Such a process may involve DNA gyrase bringing together distant PU sequences by binding them individually. If PU sequences were frequent sites of recombination, however, a much larger variety of duplications and rearrangements would be expected. Perhaps this is regulated, in part, by HU, which inhibits cruciform formation.

Overview

The multiple BIME structures, already shown to have different affinities for DNA gyrase and IHF, may also interact differentially with other BIME-associated proteins. The combination of motifs present within a particular BIME could influence its function, and each biochemical property associated with BIMEs should be studied with regard to the existence of the different families. Since PU sequences in *E.coli* are not usually found at corresponding loci in *Salmonella typhimurium*, rigidly preserved positions may be unnecessary for most, if not all BIME functions. This suggests a widespread rather than localised roles.

1.10 Osmotic shock

E.coli is able to survive outside its preferred habitat in the gut of vertebrate animals, and has developed protective systems against various stresses in order to do so. Such stresses include high osmolarity arising under conditions of desiccation or upon exposure to seawater. In hyperosmotic media, bacteria must increase their internal solute concentrations to avoid plasmolysis and maintain turgor pressure.The primary response of *E.coli* to osmotic stress is a rapid accumulation of K^+ ions, achieved in part through the high-affinity potassium transport system, the product of the *kdp* operon. Transcription of *kdp* is induced by limited availability of K^+ even at low osmolarity, and repressed at high osmolarity on addition of K^+ (Sutherland *et al.*, 1986). The stimulus for osmotic induction of *kdp* is unclear, but it may not be directly linked to the loss of turgor pressure as originally suggested (Laimins *et al.*, 1981). Asha and Gowrishankar (1993) found that while several ionic solutes were able to induce *kdp* expression, non-ionic solutes were not, even though they also increased osmolarity. In addition, intracellular solutes including trehalose and glycine betaine, which would be expected to maintain turgor pressure, had no effect on *kdp* expression. The authors proposed that the signal is related to the K^+ flux across the membrane during shock. The intracellular glutamate concentration, like that of K^+, increases in proportion to the osmolarity of the growth medium, contributing to the maintenance of electrical neutrality (Richey *et al.*, 1987).

Osmotic protectants such as trehalose, choline, and glycine betaine are accumulated as a secondary response to osmotic stress. High intracellular concentrations of salts would seriously affect metabolic function and maintenance of membrane potentials, and osmolyte systems have developed in many organisms to be compatible with macromolecules (Yancey *et al.*, 1982). These compatible solutes serve as osmoprotectants by maintaining both cell turgor pressure and intracellular enzyme function, and significantly enhance growth rates at high osmolarity. Glycine

betaine is the preferred osmoprotectant in *E.coli* (Sutherland *et al.,* 1986) and is able to protect proteins from ionic or other denaturation (Yancey *et al.,* 1982).

Uptake of glycine betaine is mediated by two distinct transport systems encoded by *proP* and *proU* (Gowrishankar, 1985; May *et al.,* 1986; Stirling *et al.,* 1989). ProP mediates relatively low-affinity transport (Cairney *et al.,* 1985a), whereas ProU specifies a high-affinity transport system whose synthesis is induced only when cells are grown at high osmolarity (Cairney *et al.,* 1985b; Gowrishankar,1985; May *et al.,* 1986). *proU* expression *in vivo* is induced about 100-fold by elevated osmolarity (May *et al.,* 1989) and appears to require increased K^+ concentrations (Sutherland *et al.,* 1986). The increase in *proU* expression is mediated, at least in part, by changes in the DNA supercoiling level (Higgins *et al.,* 1988a), while addition of glycine betaine prevents induction by osmotic shock (Sutherland *et al.,* 1986).

1.10.1 Induction of Osmotic shock genes

Exposure to high osmolarity induces at least 20 proteins as determined by two-dimensional gel electrophoresis (Botsford, 1990; Hengge-Aronis *et al.,* 1993; Jenkins *et al.,* 1990). The majority of these proteins are found in the cell envelope, but little is known about their functions and how they might aid the cell in coping with osmotic stress. However, blocking expression of osmotically-induced proteins has demonstrated their importance to hyperosmotic stress protection (Jenkins *et al.,* 1990).

Many osmotically-inducible genes have been identified by transposon mutagenesis-created reporter gene fusions. These include *osmB* and *osmY. osmB* encodes a lipoprotein located in the outer membrane (Jung *et al.,* 1989) which bears a strong structural similarity to the major *E.coli* lipoprotein, the product of *lpp* (Braun, 1975; Inouye *et al.,* 1977). *osmY* encodes an osmotically-inducible periplasmic protein (Yim and Villarejo, 1992). This gene is identical to the carbon-

starvation-inducible gene *csi-5* and is under the control of *rpoS* (Hengge-Aronis *et al.,* 1993). *rpoS* is discussed in the next section.

Genetic loci responsive to hyperosmotic stress are distributed throughout the entire *E.coli* chromosome (Csonka, 1989; Yim and Villarejo, 1992) but as yet, no central regulatory protein for the stimulon has been found. Some of the loci within the stimulon are known to be controlled by members of a family of specific regulatory factors, whose members in general control responses to environmental stimuli and are related by amino acid sequence homology (Ronson *et al.,* 1987). In these two-component systems found in many species of bacteria, a sensor protein detects the environmental stimulus and transmits a signal to an effector or regulator component that elicits a response, usually at the level of transcription. Phosphorylation of the smaller cytoplasmic effector by the larger membrane-bound sensor has been shown to occur for example in the control of porin synthesis (Forst *et al.,* 1989), nitrogen regulation (Magasanik, 1988; Ninfa and Bennet, 1991), and chemotaxis (Hess *et al.,* 1988; Wylie *et al.,* 1988). In the first of these examples, genes encoding the osmotically inducible outer membrane porins OmpC and OmpF are under the control of the osmotic sensor EnvZ and the transcriptional regulator OmpR. Regulation of the *kdp* operon similarly requires KdpD, a 99-kD membrane protein, and KdpE, a 25-kD cytoplasmic protein. Their amino acid sequences show that they are members of this same sensor-effector family (Walderhaug *et al.,* 1992).

In a further example, the *cps* genes involved in colanic acid capsule synthesis are regulated by RcsB, the effector, and RcsC, the sensor (Stout and Gottesman, 1990). Colanic acid is a mucoid exopolysaccharide that is normally synthesised by *E.coli* K12 in low amounts. Certain mutations that affect lipopolysaccharide synthesis and structure, leading to changes in the outer membrane, can increase colanic acid synthesis via activation of the inner membrane-bound RcsC, which in turn activates RcsB (Parker *et al.,* 1992). This suggests that one function for colanic acid might be to protect the *E.coli* outer membrane from environmental stress. It has indeed been shown that colanic acid protects *E.coli* from dessication (Ophir and Gutnick, 1994) and recently Sledjeski and Gottesman (1996) found that a *cps*::*lacZ* fusion was

osmotically inducible. An alternative pathway for activation of *cps* gene transcription involves the unstable positive regulator RcsA. This is normally unstable because it is a substrate of Lon protease. If sufficient RcsA accumulates, *cps* transcription is activated independently of RcsC. Genetic and biochemical data suggest that RcsA and RcsB can interact to activate *cps* transcription (Brill *et al.,* 1988; Stout *et al.,* 1991). Overproduction of capsular polysaccharide in *lon* mutant hosts is probably due to RcsA stabilization (Torres-Cabassa and Gottesman, 1987). Osmotic induction of *cps* was found to be dependent on the RcsB-RcsC pathway, with RcsA playing only a stimulatory role (Sledjeski and Gottesman, 1996).

Although these three similar regulatory pathways respond to the same stimulus, osmotic shock, their signal transduction pathways appear to be specific. For example, the RcsC-RcsB system is similar to the KdpD-KdpE system, but the finding that RcsC is essential for osmotic induction of *cps* indicates that stimulated KdpD cannot cross-talk detectably to RcsB. This suggests that these systems are geared to different degrees of osmotic shock. For example, a slight increase in osmolarity may require activation of the *kdp* operon to uptake K^+, but be too small a stimulus to warrant capsule synthesis. The systems may also respond selectively to alternate stimuli, with *cps* induction occuring preferentially under conditions that stabilize RcsA. If each osmotically-induced operon is individually controlled, there may not be a central regulatory protein.

1.10.2 *rpoS* and osmotic shock

Several osmotically-inducible genes such as *osmB* and *osmY* are also induced when cells enter stationary phase (Jenkins *et al.,* 1990). This induction is believed to depend on *rpoS*, whose product is the "stationary phase" transcription initiation subunit, σ^S. However, the intracellular level of σ^S does not significantly increase when exponentially-growing *E.coli* cells are exposed to osmotic shock, and many other *rpoS*-dependent genes are not osmotically inducible (Hengge-Aronis *et al.,*

1993). Thus, within the *rpoS* regulon there exists a subset of genes which are also induced by increased osmolarity. Presumably the induction of *osmB* and *osmY* involves some activator which can stimulate RNA polymerase containing σ^S, but not the major exponential phase sigma, σ^{70}.

Yim *et al.* (1994) found that *osmY* is regulated at the transcriptional level and that a single promoter is responsive to both osmotic and growth-arrest stimuli. In contrast, Jung *et al.* (1990) found two transcription control sites of *osmB.* Transcription from the downstream promoter P2 was induced by elevated osmolarity or upon entry into stationary phase, whereas transcription from P1, 150bp upstream, occured only when both osmotic and growth phase signals arose simultaneously. Deletion analysis identified a 7-bp sequence just upstream of the -35 element of P2 essential for osmotic stimulation, but not for growth-arrest induction.

Using *lacZ* fusions, Yim *et al.* (1994) showed that *osmY* and *rpoS* fusions were expressed in parallel, and that *osmY* expression did not require a functional σ^{70}, suggesting a direct role for σ^S. Thus, *osmY* may be induced as a consequence of the 5-10 fold increase in σ^S level when cells approach stationary phase (Tanaka *et al.,* 1993). Following osmotic shock, *osmY* may be activated by changes in K^+ levels or supercoiling without the need to increase σ^S concentration (although σ^S is still required). In addition, core RNA polymerase and/or σ^S could be modified during hyperosmotic stress so that the affinity for certain promoters is increased. Transcription from *osmB* P1 may be regulated by a different mechanism.

To further study transcription of *osmY* and *osmB,* Ding *et al.* (1995) used *in vitro* transcription assays directed by reconstituted RNA polymerase holoenzyme containing either σ^{70} or σ^S. They found that the relative levels of transcription by the two enzymes varied depending on reaction conditions. At low ionic strength, the *osmB* and *osmY* promoters were transcribed by both $E\sigma^{70}$ and $E\sigma^S$. However, addition of up to 400mM potassium glutamate, to mimic intracellular conditions under hyperosmotic stress, specifically enhanced transcription from both promoters by $E\sigma^S$, while inhibiting that by $E\sigma^{70}$. Similar high concentrations of potassium chloride inhibit both forms of RNA polymerase, suggesting that $E\sigma^S$ itself can sense

osmotic stress by responding to changes in intracellular potassium glutamate concentration resulting in increased affinity for osmoregulated promoters.*In vivo*, however, potassium glutamate may act to relieve the transcriptional repression caused by binding of H-NS (see next section).

The effects of supercoiling on transcription from the *osmY* promoter were investigated by Kusano *et al.* (1996). They found that maximum transcription of DNA with low superhelicity required lower concentrations of potassium glutamate than that of DNA with high superhelicity. DNA superhelicity is known to change depending on cell growth conditions. For instance, nutrient downshift and entry into stationary phase cause a decrease in DNA superhelicity (Balke and Gralla, 1987; Jaworski *et al.*, 1991) while high osmolarity leads to an increase (Higgins *et al.*, 1988a; Hsieh *et al.*, 1991).

OsmY and OsmB may play the same protective role in stationary phase growth and hyperosmotic conditions by altering the structure of the periplasm and outer membrane. *osmY* transcription is induced on entry into stationary phase by the increase in the concentration of σ^S and the decrease in negative supercoiling, while the stimulus following osmotic shock is the increase in potassium glutamate levels. Within the *rpoS* regulon it seems likely that each of the σ^S-dependent promoters requires specific factors or growth conditions for efficient transcription.

1.10.3 H-NS and osmotic shock

As mentioned in section 1.8.1, H-NS plays a regulatory role in determining the concentration of σ^S, and may thus indirectly affect transcription of σ^S-dependent osmotically induced genes. H-NS also plays a direct role in the transcriptional control of other osmotically induced genes in a σ^S-independent manner.

Ueguchi and Mizuno (1993) showed that purified H-NS inhibited transcription from the *proU* promoter *in vitro*, and DNaseI footprinting provided evidence of ~150bp protection upstream of the -35 promoter element. The inhibitory effect was

found to be alleviated by increased superhelicity of the template DNA, and strikingly more so by increased concentrations of K^+, in particular potassium glutamate. H-NS repressed the *proU* promoter specifically, since it had no effect on transcription from an upstream *tac* promoter. Binding of H-NS downstream of the promoter, on the DNA encoding the 5′ end of the *proU* mRNA has also been demonstrated (Lucht *et al.*, 1994; Mellies *et al.*, 1994). There is evidence of *proU* derepression in mutants lacking this negative downstream regulatory element.

Trans-acting regulatory mutants with reduced *proU* expression were found to be allelic to *rpoS* and *hupB*, the latter encoding the B subunit of the histone-like protein HU (Manna and Gowrishankar, 1994). Thus HU plays a positive role in *proU* regulation in contrast to H-NS. HU also plays a positive role in *rcsA* expression, and HU oversynthesis induces mucoidy in *E.coli* (Painbeni *et al.*, 1993).

The *otsBA* operon encoding the trehalose synthesis system was found to be more strongly expressed in a *hns rpoS* double mutant than in a mutant deficient for *rpoS* alone, indicating that H-NS plays a direct negative role in *otsBA* expression (Barth *et al.*, 1995). In general the direct H-NS effect can have varying importance in the regulation of σ^S-dependent genes. In the case of *csgBA* high levels of expression were seen in a *hns rpoS* double mutant (Arnqvist *et al.*, 1994), indicating that the repression by H-NS is of primary importance. The *csg* operon encodes curli, which are fibronectin and laminin-binding fibres expressed by *E.coli* strains in response to low temperature, low osmolarity, and stationary growth conditions. On the other hand, a mutation in *hns* could only partially suppress the requirement for σ^S in the expression of *otsBA* (Barth *et al.*, 1995).

The normally low level of transcription from the promoter of *rcsA*, which encodes the positive regulator of colanic acid synthesis, is thought to be due to H-NS repression. Using a *lacZ* fusion, Sledjeski and Gottesman (1995) found that *rcsA* transcription was stimulated following overproduction of DsrA, an 85nt RNA whose coding region lies just downstream of *rcsA*. The transcription start site was found to be the same with or without DsrA, showing that this RNA does not mediate transcription from another promoter. Increased transcription was also found when the

region upstream of -59 of the *rcsA* promoter was deleted, or when a mutant H-NS was present. In these strains, DsrA had no effect on *rcsA* transcription, suggesting that rather than behaving as a direct activator, DsrA overcomes the silencing action of H-NS. DsrA has no sequence homology either with the *rcsA* promoter or with any other sequenced *E.coli* DNA, although it is predicted to form an identical secondary structure to that of the homologue in *Klebsiella pneumoniae,* with which it has 84% sequence identity. The effect of DsrA was lost when either the promoter or the terminator of *dsrA* transcription was removed, suggesting that RNA secondary structure is important. The antisilencing effect was also seen when just the *rcsA* promoter and upstream region was fused to a *lacZ* reporter gene, showing that no sequences within the *rcsA* transcribed region are important. The RNA may interfere with H-NS binding to the promoter of *rcsA,* or prevent H-NS from forming a chromatin-like silencing structure. This is the only RNA molecule known to play a positive role in transcription, although others may exist. Transcription of antisilencing genes may be activated following a particular stimulus, such as increased osmolarity or entry into stationary phase.

The complex regulation of genes expressed following osmotic shock may indicate a specific requirement for their protein products to permit cell growth also under other stressful conditions. This may also explain why there is no single regulator, such as an alternative sigma factor, to direct osmotic shock gene transcription.

1.11 The Cold-shock response

Bacteria are well adapted to respond to instantaneous changes in temperature, with the induction of a specific set of genes upon temperature shift. Jones *et al.* (1987) found that transferring *E.coli* from 37°C to 10°C results in a 4 hour lag followed by resumption of growth with a generation time of 24 hours. During the lag period most protein synthesis ceases, except for a specific set of cold-shock-induced proteins. Generally, the magnitude of induction corresponds to the range of temperature downshift, and it is believed that the cold-shock response serves an adaptive function for continued growth at low temperatures.

1.11.1 Cold-shock proteins in *E.coli*

Most cold-shock proteins are synthesised at differential rates 2-10 times greater than their rates at 37°C (Jones *et al.,* 1987). The identified cold-shock proteins are involved in various cellular processes such as transcription (NusA), translation (InfB), mRNA degradation (PnPase), recombination (RecA), and DNA topology (H-NS and GyrA).

The most striking induction is that of CspA, whose concentration increases 200-fold following temperature downshift from 37°C to 10°C. As such, CspA has been designated the major cold-shock protein (Jones *et al.,* 1987). The rate of CspA production accounts for 13% of total protein synthesis within 90 minutes at 10°C, and subsequently drops to a lower basal level (Goldstein *et al.,* 1990). Induction of CspA was originally thought to arise primarily at the level of transcription, involving the binding of an unknown factor upstream of the *cspA* promoter (Tanabe *et al.*, 1992). However, Goldenberg *et al.* (1996) demonstrated that the half life of *cspA* mRNA increased considerably upon downshift to 15°C. In addition, they found that the *cspA* promoter fused to *lacZ* did not significantly increase β-galactosidase activity at 15°C, indicating that posttranscriptional mechanisms must play the major role in CspA regulation. Brandi *et al.* (1996) showed that cell-free S30 extracts from

cold-shocked cells were able to translate *cspA* mRNA much more efficiently than those from non-shocked cells, while both extracts translated a control transcript with equal efficiency. They proposed that the translational machinery is modified during cold-shock so as to preferentially translate this transcript. Although the mechanisms involved in this selective translation are unknown, they may also play a part in increased *cspA* transcript stability due to the occupation of mRNA by ribosomes.

1.11.2 Role of CspA in cold-shock

CspA is a 70 aa protein with high sequence similarity to one domain of eucaryotic Y-box transcription factors (Wistow, 1990; Wolffe *et al.*, 1992). These proteins recognise a specific regulatory sequence, the Y-box, which contains a conserved ATTGG motif (Wolffe, 1994). CspA has been implicated in the transcriptional regulation of two cold-shock genes, *hns* and *gyrA*. The regulatory domain of *hns* contains an ATTGG motif, and CspA binding to the *hns* promoter has been demonstrated *in vitro* (La Teana *et al.*, 1991). CspA has also been shown to bind to the promoter region of *gyrA*, and mutation of three ATTGG motifs in this region eliminated specific protein-DNA interaction (Jones *et al.*, 1992b). CspA may assist transcription initiation by facilitating open complex formation at low temperatures, perhaps by binding specifically to the non-coding strand around the transcription start site.

It has been suggested that CspA might also bind RNA molecules, and may be involved in the translation of tightly folded mRNA molecules. Jones *et al.* (1996) identified a cold-shock-inducible ribosome-associated protein which was found to unwind double stranded RNA. They proposed that this 70kD protein, CsdA, plays an essential role in translation by unwinding cold-stabilized secondary mRNA structures. Since CspA is a putative 'RNA chaperone' (Jones and Inouye, 1994), it may bind to mRNA unwound by CsdA so as to prevent reannealing.

In addition to CspA, *E.coli* contains a family of CspA-like proteins among which only CspB is cold-inducible (Jones and Inouye, 1994; Yamanaka *et al.,* 1994). Although the functions of these proteins are unknown, it has been suggested that they may play a role in masking the translation of certain mRNA molecules by recognising specific RNA structures (Jones and Inouye, 1994).

1.11.3 Regulation of cold-shock gene expression

DNA gyrase-mediated supercoiling appears to be involved in the regulation of cold-shock gene expression since induction was reported to have been lost in a gyrase deficient strain (Qoronfleh *et al.,* 1992). In addition, the DNA binding protein H-NS is thought to play an important role in the adaptation to low temperatures since Dersch *et al.* (1994) found growth inhibition in *hns* mutants. It seems likely that CspA induction is the primary response to cold-shock, leading to a cascade of other effects. Induction of *gyrA* and *hns* transcription, due to CspA binding at their promoters, results in other cold-shock gene expression via DNA topology. In addition, it is possible that CspA directly affects transcription of certain cold-shock genes.

Other possible participants in the regulatory process are guanosine 5′triphosphate-3′ diphosphate and guanosine 5′diphosphate-3′diphosphate (collectively (p)ppGpp). Levels of (p)ppGpp decrease in proportion to the extent of temperature downshift (Pao and Dyess, 1981; Mackow and Chang, 1983). Artificially increasing (p)ppGpp levels prior to temperature downshift resulted in decreased induction of cold-shock proteins, while a *relA spoT* mutant (which lacks detectable levels of (p)ppGpp) produced higher levels of cold-shock-induced proteins (Jones *et al.,* 1992a). It has been suggested that following temperature downshift the decreased translational capacity of the cell relative to the supply of charged tRNA triggers the decrease in (p)ppGpp levels, resulting in some of the altered gene expression observed.

The addition of certain inhibitors of translation (e.g. chloramphenicol, tetracycline and fusidic acid) was also found to induce a cold-shock-like response, leading to the proposal that the state of the ribosome is the normal physiological sensor for induction of this response (VanBogelen and Neidhardt, 1990). Many translational inhibitors that induce the cold-shock response also decrease the level of (p)ppGpp (Lund and Kjeldgaard, 1972), consistent with the proposed central involvement of these nucleotides in the cold-shock response.

The function of the cold-shock response is not known, but the specific pattern of gene expression indicates that it is probably required for optimal adaptation to growth at low temperatures. In *E.coli*, a temperature downshift appears to modify the translational machinery leading to preferential synthesis of cold-shock proteins. In addition to stimulation of transcription, increased stability of cold-shock mRNAs and proteins may lead to their elevated levels at lower temperatures. Although the mechanisms involved are unclear, continued study of the cold-shock response will provide valuable insight into how cells respond and adapt to various stresses.

1.12 Aromatic Catabolism

Studies of bacterial degradation of aromatic compounds have focussed mainly on the genus *Pseudomonas* and other soil bacteria. The initial pathway for the catabolism of various substrates including benzene, toluene, xylene, phenol and biphenyls is carried out by specific enzymes which convert the compounds to catecholic (1,2-dihydroxybenzene) intermediates. However, the reactions involving oxygenative ring fission of the catechol and the subsequent conversion to Krebs cycle intermediates are limited to one of two alternatives. The *ortho*-cleavage pathways involve ring cleavage between the two hydroxyl groups followed by a well defined series of reactions leading to β-ketoadipate (reviewed by Dagley, 1986). The *meta*-cleavage pathway involves ring cleavage adjacent to the two catechol hydroxyls, followed by degradation of the ring cleavage product to pyruvate and a short chain aldehyde (see fig. 1.5). The enzymes involved in these pathways catalyse the same reactions but are specific to a particular substrate.

Meta-cleavage pathways can metabolize many substituted aromatic compounds (Worsey and Williams, 1975), which may reflect the availability of different intermediates in nature. The *tod* gene cluster of *Pseudomonas putida* F1 encodes enzymes for toluene degradation (Lau *et al.,* 1994), while the TOL plasmid pWWO of *P.putida* mt-2 encodes enzymes required for the oxidation of toluene and xylenes (Harayama and Rekik, 1993). Phenol and its methyl derivatives are degraded by the plasmid pVI150-encoded *dmp* gene products of *Pseudomonas* sp. strain CF600 (Shingler *et al.,* 1992) while the *bph* loci of *Pseudomonas* sp. LB400 and KK5102 allow these strains to degrade biphenyls and polychlorinated biphenyls (Hofer *et al.,* 1994; Kikuchi *et al.,* 1994). The genes encoding the *meta*-cleavage pathway enzymes in each of these strains are present within a single operon. Comparison of the order and nucleotide sequences of these genes indicates that an ancestral gene cluster has spread horizontally and has been integrated into several operons that degrade different aromatic compounds (Harayama and Rekik, 1993, and references therein).

1.12.1 Aromatic catabolism in *E.coli*

Several species of enteric bacteria including *E.coli* are also known to posess aromatic degradation pathways. Cooper and Skinner (1980) first demonstrated this by showing that strains B, C and W of *E.coli* (but not K12) were able to grow on 3- and 4-hydroxyphenylacetate , and that all the enzymes necessary were present in cell extracts prepared from induced cells. The authors proposed that the strains of *E.coli* that they used had evolved the ability to grow on 4-hydroxyphenylacetate because this compound, being a product of anaerobic degradation of tyrosine, would be encountered in faecal environments.

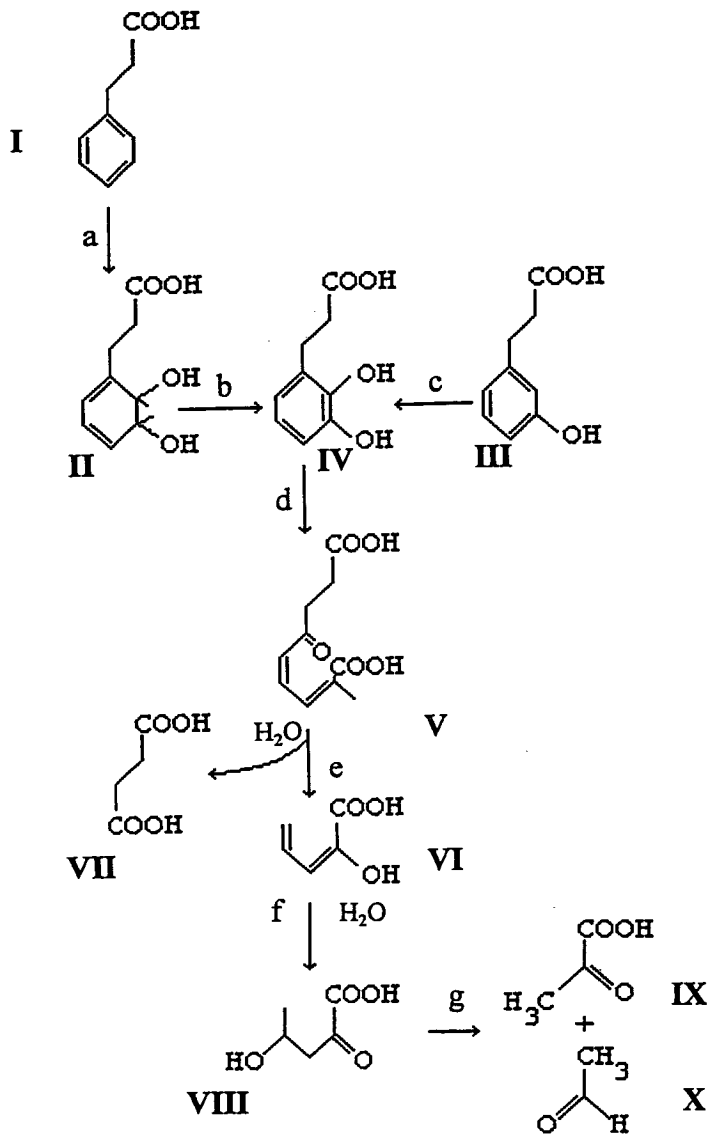
Burlingame and Chapman (1983) went on to demonstrate that *E.coli* K12, as well as strains B, C, and W, were able to utilize 3-phenylpropionic and 3-(3-hydroxyphenyl)propionic acids as sole sources of carbon for growth. Fig. 1.5 shows their proposed reaction sequence for the degradation of 3-phenylpropionic acid and 3-(3-hydroxyphenyl)propionic acid to succinate, acetaldehyde and pyruvate. These substrates are also compounds that have been found in the urine and faeces of animals, and are formed by the action of intestinal microflora on various plant constituents (Burlingame and Chapman, 1983, and references therein). The genes encoding the necessary enzymes have not yet been identified.

Burlingame and Chapman (1983) also demonstrated the ability of strains K12 and W to utilize phenylacetic acid as carbon source. Cooper *et al.* (1985) identified K12 mutations which mapped to 30.4 minutes on the genetic map and which prevented growth on this acid. Later Parrott *et al.* (1987) showed that *E.coli* K12 can grow on 2-phenylethylamine, and proposed that this compound is first converted to phenylacetic acid. There is still uncertainty regarding the pathway for phenylacetate metabolism, but analysis of the DNA sequence around 30.4 minutes, when complete, may help to resolve this.

Figure 1.5. Proposed reaction sequence for the degradation of 3-phenylpropionic acid and 3-(3-hydroxyphenyl)propionic acid by *E.coli* strains K12, B, C, and W (from Burlingame and Chapman, 1983).

The metabolites are as follows: 3-phenylpropionic acid (I); *cis*-3-(3-carboxyethyl)-3,5-cyclohexadiene-1,2-diol (II); 3-(3-hydroxyphenyl)propionic acid (III); 3-(2,3-dihydroxyphenyl)propionic acid (IV); 2-hydroxy-6-ketona-2,4-dienedioic acid (V); 2-keto-4-pentenoic acid (enol) (VI); succinic acid (VII); 4-hydroxy-2-oxovalerate (VIII); acetaldehyde (IX); pyruvate (X).

The enzymes catalysing each step are as follows: 3-phenylpropionic acid oxygenase (a); *cis*-3-(3-carboxyethyl)-3,5-cyclohexadiene-1,2-diol isomerase (b); 3-(3-hydroxyphenyl)propionic acid hydroxylase (c); 3-(2,3-dihydroxyphenyl)propionic acid dioxygenase (d); 2-hydroxy-6-ketona-2,4-dienedioic acid hydrolase (e); 2-keto-4-pentenoic acid hydrolase (f); 4-hydroxy-2-oxovalerate aldolase (g).

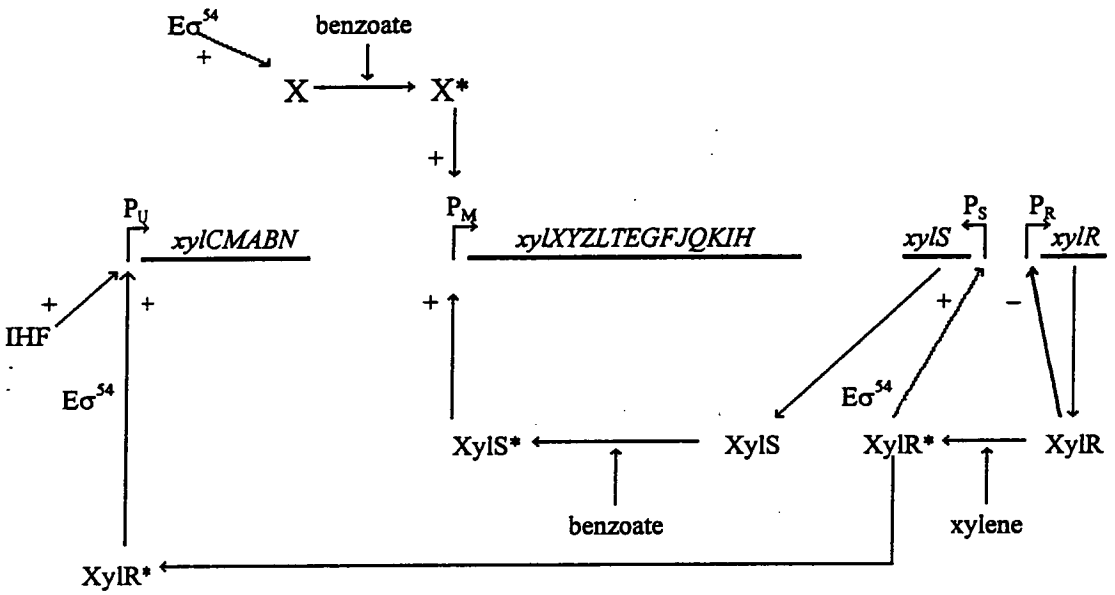


Jenkins and Cooper (1988) published the first study of the homoprotocatechuate (3,4-dihydroxyphenylacetate) degradative pathway in *E.coli* C, while more recently Stringfellow *et al.* (1995) completed the sequence of the *hpc* operon involved. The *meta*-cleavage pathways for catechol and homoprotocatechuate degradation involve chemically analogous reactions, but *E.coli* K12 strains appear to be devoid of *hpc* genes, and do not catabolise homoprotocatechuate (Jenkins and Cooper, 1988). Roper *et al.* (1993) noticed that an open reading frame just upstream of the *hpc* operon in *E.*coli C shared 99% identity with the *E.coli* K12 *tsr* gene encoding a serine receptor protein. *tsr* maps to 98.9 minutes on the K12 physical map, but the location of *tsr* on the *E.coli* C map has not been reported.

1.12.2 Regulation of *xyl* gene expression in *Pseudomonas*

The transcriptional control of the *xyl* operon carried by the TOL plasmid pWWO has been extensively studied (reviewed by Marqués and Ramos, 1993). Expression of the upper pathway operon, which encodes enzymes for the oxidation of toluene/xylenes to benzoate/toluates, and the subsequent *meta*-cleavage pathway operon, is controlled by the products of the regulatory genes *xylR* and *xylS.* These genes are located adjacent to each other at the downstream end of the *meta*-cleavage operon, and are transcribed from divergent promoters. XylR stimulates transcription from the upper pathway operon promoter (Pu) in the presence of pathway substrates, and also from the *xylS* promoter (Ps). Both promoters are σ^{54}-dependent, and transcription from Pu also requires Integration Host Factor. XylS-stimulated transcription from the σ^{70}-dependent *meta*-cleavage pathway operon promoter (Pm) is strongly enhanced by the presence of substrate benzoates. Expression of the *meta*-cleavage operon is also induced by xylene-activated XylR which additionally stimulates transcription from Ps independently of *meta*-cleavage pathway substrates such as benzoate. The overall regulation of the *xyl* operons is shown in fig.1.6.

Figure 1.6. Regulation of the *xyl* operons in *Pseudomonas*



- Positive regulation of the upper pathway operon at P_U by $E\sigma^{54}$/XylR* and IHF
- Positive regulation of the *meta*-cleavage operon at P_M by $E\sigma^{70}$/XylS* and unknown activator, X.
- Positive regulation at P_S by $E\sigma^{54}$/XylR*.
- Autoregulation at P_R by $E\sigma^{70}$/XylR.

Genes are represented by bold lines with the corresponding names and gene orders within the two operons given above the line. The arrows at the four promoters P_U, P_M, P_S, and P_R indicate the direction of transcription. XylS* and XylR* are the pathway substrate-activated forms of XylS and XylR respectively. + indicates positive regulation of transcription; - indicates autoregulation at P_R.

The C-terminal end of XylS is highly homologous to at least 27 proteins involved in transcriptional stimulation in carbon metabolism, pathogenesis and in response to alkylating agents (Gallegos *et al.*, 1993; Ramos *et al.*, 1990). A conserved helix-turn-helix motif in this region is probably involved in DNA binding. This set of regulators constitutes the AraC/XylS family. The interaction of benzoates with XylS leading to activation of the regulator is poorly understood, although amino acids in the N-terminus have been implicated in effector binding. One possibility is that an intramolecular signal transmitted from the N-terminus to the C-terminus exposes the DNA binding motif so as to activate transcription.

Kessler *et al.* (1994) noted XylS-independent induction of transcription from the Pm promoter by benzoate. This did not occur in an *rpoN* mutant background, suggesting that a σ^{54}-dependent regulator (X in fig. 1.6) can activate transcription of the *meta*-cleavage operon in the absence of XylS.

The XylR protein belongs to the NtrC family (Inouye *et al.*, 1988) and shares 67% overall identity with DmpR, the positive regulator of the phenol catabolic pathway borne by the *Pseudomonas* plasmid pVI150 (Shingler *et al.*, 1993). These regulators exhibit four domains, three of which are highly conserved within the family. The C-terminal domain carries a helix-turn-helix DNA binding motif, while the non-conserved N-terminal domain seems to be involved with signal interaction in the case of DmpR (Shingler *et al.*, 1993) and XylR (Abril *et al.*, 1989; Inouye *et al.*, 1988). XylR is constitutively expressed and autoregulates its own synthesis (Abril *et al.*, 1989), but only stimulates transcription from Pu and Ps in the presence of pathway substrates or effectors. Using non-metabolizable effectors of both XylR and XylS, Duetz *et al.* (1994) demonstrated catabolite repression of the upper and *meta*-cleavage operons by succinate.

1.12.3 Control of aromatic catabolic gene expression in *E.coli*

Although positive regulation governs the expression of many aromatic catabolic pathways, the homoprotocatechuate (3,4-dihydroxyphenylacetate) degradative operon of *E.coli* C appears to be negatively regulated (Jenkins and Cooper, 1988). Removal of *hpcR,* the gene encoding the putative *hpc* repressor, led to constitutive expression of the *hpc* operon. *hpcR* was found to lie upstream of a promoter/operator region of *hpc*. Its sequence showed no significant similarity with other regulatory genes (Roper *et al.,* 1993). The putative operator was found to contain two inverted repeats separated by 55bp, which may represent a binding site for HpcR. In addition, a binding site for the Catabolite Activator Protein was centred 20bp upstream of the *hpc* promoter, consistent with the reported glucose repression of this pathway (Jenkins, 1987).

While investigating the catabolism of 3- and 4-hydroxyphenylacetate, Cooper and Skinner (1980) showed that the necessary enzymes were present in cell extracts prepared from induced cells, but were absent from non-induced cells. The regulation of this pathway is still unclear, but sequencing and identification of the genes involved may reveal similarities with other operons.

Considering the similarities of the genes involved in aromatic degradation in all the bacteria studied, it seems likely that common mechanisms control their expression. Catabolite repression in the presence of more favourable substrates and induction of the necessary genes when aromatic compounds are the only metabolites available appear to be common features.

Chapter 2

Materials and Methods

2.1 MATERIALS

2.1.1 Chemicals and enzymes

Most chemicals were of analytical or molecular biology grade and were bought from the following suppliers (unless otherwise stated) :

Aldrich Chemical Company Ltd., Gillingham, Dorset.

BDH Ltd., Poole, Dorset.

Boehringer Corporation (London) Ltd., Lewes, Sussex.

Fisons plc, Loughborough, Leicestershire.

Sigma Chemical Company, Poole, Dorset.

Agarose for gel electrophoresis of DNA and RNA was obtained from FMC Corporation, Rockland, ME, USA. Low melting point agarose was obtained from BRL (UK) Ltd., Cambridge.

Hybond N^+ positively charged nylon membrane, linked T7 Transcription-Translation system and radiolabelled compounds were obtained from Amersham International plc, Amersham, Bucks.

Enzymes used for nucleic acid manipulation purposes were purchased from :

Boehringer Corporation (London) Ltd., Lewes, Sussex.

New England Biolabs, Beverly, MA, USA.

Pharmacia LKB Biotechnology, Milton Keynes, Bucks.

United States Biochemical Corporation, Cleveland, Ohio, USA.

Synthetic oligonucleotides were purchased from the Oswel DNA Service of Edinburgh University.

Purified RNA polymerase core enzyme, σ^{70}, and σ^{S} were kind gifts from Professor Akira Ishihama, National Institute of Genetics, Mishima, Japan.

2.1.2 Standard solutions

Solutions were made up with dH_2O and sterilised by autoclaving at 120°C, 15psi for 15 minutes, or by filtration through Acrodisc filters (0.45μm pore size, Gelman Sciences).

Solutions were stored at room temperature, unless otherwise stated.

1 x TE	10mM Tris-HCl, 1mM EDTA; adjusted to pH 8.0 with concentrated HCl.
10 x TAE	0.4M Tris base, 0.2M sodium acetate, 10mM EDTA; adjusted to pH 8.3 with glacial acetic acid.
TAE Agarose Gel Loading Buffer	10ml 10 x TAE, 3g Ficoll, 1ml 10% (w/v) bromophenol blue in methanol.
10 x TBE	0.89M Tris base, 0.89M boric acid, 20mM EDTA; pH 8.2 without adjustment.
TBE Agarose Gel Loading Buffer	10ml 10 x TBE, 3g Ficoll, 1ml 10% (w/v) bromophenol blue in methanol.
20 x SSC	3.0M NaCl, 0.3M trisodium citrate; adjusted to pH 7.0.
50 x Denhardt's solution	1% (w/v) Ficoll, 1% (w/v) polyvinylpyrollidine, 1% (w/v) BSA (Pentax fraction V).
10 x Tris-glycine Running Buffer	3% (w/v) Tris base. 14.4% (w/v) glycine; adjusted to pH 8.6.
RNase A	5mg/ml. Heated to 100°C for 15 minutes, then cooled to room temperature before storage, in aliquots, at -20°C.
Phenol	Obtained from Sigma Chemical Company. Saturated with 1 x TE buffer for extractions of DNA solutions; saturated with 0.1M citrate buffer, pH 4.3, for extractions of RNA solutions.
Chloroform	Chloroform and isoamylalcohol mixed 24:1 (w/v); stored in darkened bottles at room temperature.

IPTG	25mg/ml stored at -20°C.
X-Gal	25mg/ml in dimethyl formamide. Stored at -20°C.

2.1.3. Growth media

Growth media were made up with dH_2O and sterilised by autoclaving or by filter sterilisation. Quantities listed are for 1 litre final volume, unless otherwise stated.

L-broth (LB)	10g Difco bactotryptone, 5g bacto yeast extract, 10g NaCl; adjusted to pH 7.2 with concentrated NaOH.
LB Agar	LB plus 15g Difco agar.
BBL Agar	10g Baltimore Biological Laboratories trypticase, 5g NaCl, 10g Difco agar.
BBL Top Agar	As for BBL Agar, but with 6.5g Difco agar.
Freezing Broth	40g Difco nutrient broth, 5g NaCl.
4 x M9 Salts	28g Na_2HPO_4, 12g KH_2PO_4, 2g NaCl, 4g NH_4Cl.
M9 Minimal Medium	100ml 4 x M9 salts, 300ml dH_2O, plus 0.2% (w/v) sugars, 20μg/ml amino acids, 2μg/ml thiamine, $MgSO_4$ (80mg).
Spizizen Minimal Medium	2g $(NH_4)_2SO_4$, 14g K_2HPO_4, 6g KH_2PO_4, 1g trisodium citrate dihydrate, 0.2g $MgSO_4.7H_2O$, 0.2% (w/v) sugars plus 2mg thiamine, 20mg amino acids as required.
Spizizen Minimal Agar	Spizizen minimal medium plus 15g Difco agar.
Bacterial Buffer	3g KH_2PO_4, 7g Na_2HPO_4, 4g NaCl, 0.2g $MgSO_4.7H_2O$.
Phage Buffer	3g KH_2PO_4, 7g Na_2HPO_4, 5g NaCl, 10ml 0.1M $MgSO_4$, 10ml 0.1M $CaCl_2$, 1ml 1% (w/v) gelatin.

All media, with the exceptions of minimal medias and agars, were prepared by the staff of the Media Room of this Institute.

2.1.4. Antibiotic stock solutions

Ampicillin	10mg/ml in dH_2O.
Chloramphenicol	Free chloramphenicol (20mg/ml) in ethanol.
Kanamycin	20mg/ml kanamycin sulphate in dH_2O.
Rifampicin	20mg/ml in methanol (stored in the dark at 4°C for 2 weeks).

Antibiotic solutions were filter sterilised and stored in aliquots at -20°C except where stated. The required amount of stock solution was added to liquid media at room temperature and to molten agar (cooled to 55°C) immediately prior to pouring.

2.1.5 Bacterial strains

All strains were derivatives of *E.coli* K12 and are listed in table 2.1.

2.1.6 Bacteriophages

Listed in table 2.2.

2.1.7 Plasmids

Listed in table 2.3.

Table 2.1 Bacterial strains

Strain	Genotype	Reference or source
NM570	Δ(*lac-pro*) *thi supE hsdM hsdS recA*13/ F′ : *proAB*$^{+}$ *lacI*^{q} lacZΔM15	N.E. Murray (Edinburgh)
TG1	Δ(*lac-pro*) *thi supE hsdD5*/ F′ : *traD*36 *proAB*$^{+}$ *lacI*^{q} lacZΔM15	Gibson (1984)
NM621	*recD hsdR mcrB pro leu thi lac tonA*21 *supE*	N.E. Murray
P678.54	*thr leu thi minB*	K. Begg (Edinburgh)
W3110	*supE* ?	Bachman (1972)
SW100	NM621 *f229* : : *kan*	This work

Table 2.2 Bacteriophages

Bacteriophage	Description	Reference
λ139	W3110 genomic DNA from 370kb - 388kb[a] cloned between the *Bam*HI sites of λEMBL4	Kohara *et al.* (1987)
λ140	W3110 genomic DNA from 381kb - 397kb[a] cloned between the *Bam*HI sites of λEMBL4	Kohara *et al.* (1987)
λ440	W3110 genomic DNA from 2759kb - 32778kb[a] cloned between the *Bam*HI sites of λEMBL4	Kohara *et al.* (1987)
λ639	W3110 genomic DNA from 4313kb - 4327kb[a] cloned between the *Bam*HI sites of λEMBL4	Kohara *et al.* (1987)
λ640	W3110 genomic DNA from 4318kb - 4332kb[a] cloned between the *Bam*HI sites of λEMBL4	Kohara *et al.* (1987)

a. Numbers in kilobase pairs are co-ordinates on the physical map of the *E.coli* genome (Kohara *et al.*, 1987).

Table 2.2 Bacteriophages (continued)

Bacteriophage	Description	Reference or source
M13 mGP1-2	M13 mp8 carrying the T7 RNA polymerase gene 1 under the control of the *lac*UV5 promoter.	M. Khattar (Edinburgh)
M13 tg130	*lacZ*α cloning vector.	Amersham International
M13 tg131	As tg130 but with the cloning sites in the opposite orientation.	Amersham International
M13 tg130-4	3.7kb *Bam*HI/*Kpn*I fragment from λ639 cloned between the same sites in tg130.	This work
M13 tg131-3	3.7kb *Bam*HI/*Kpn*I fragment from λ639 cloned between the same sites in tg131.	This work

Table 2.3 Plasmids

Plasmid	Description	Reference or source
pBluescriptIISK	AmpR, *lacZ*α cloning vector	Stratagene Ltd. Cambridge UK
pBluescriptIIKS	As SK but with the cloning sites in the opposite orientation.	Stratagene Ltd. Cambridge UK
pET3	AmpR, vector for T7-directed transcription of inserted DNA.	Studier *et al.* (1990)
pUC4-KISS	AmpR, KanR, source of Tn*901* Kanamycin resistance cassette.	Barany (1985)
pSW3	3.7kb *Bam*HI/*Kpn*I fragment from λ639 cloned between the same sites in pBluescriptIIKS. Carries a 113bp deletion in the BIME.	This work
pSW4	3.7kb *Bam*HI/*Kpn*I fragment from λ639 cloned between the same sites in pBluescriptIIKS. Carries a 339bp deletion in the BIME.	This work

pSW5	3.6kb *Bam*HI/*Kpn*I fragment from pSW3 cloned between the same sites in tg131 to pick up additional restriction sites. 3.6kb *Hind*II/*Sma*I fragment from this M13 cloned into the *Bam*HI site of pET3.	This work
pSW6	As pSW5 but with the 3.4kb *Bam*HI/*Kpn*I fragment from pSW4 used in the initial step.	This work
pSW9	pSW4 with an internal *Stu*I deletion followed by replacement of the 1.8kb *Kpn*I/*Eco*NI fragment with a linker.	This work
pSW10	1.2kb KanR cassette from pUC4-KISS cloned into the unique *Bgl*II site of pSW9. AmpR, KanR plasmid used for the disruption of *f229*.	This work
pSW11	1.8kb *Kpn*I/*Eco*NI fragment of pSW3 replaced with a linker.	This work
pSW12	1.8kb *Bam*HI/*Kpn*I fragment of pSW11 cloned into the same sites in pBluescriptIISK.	This work
pSW13	1.3kb *Bgl*II/*Bam*HI fragment deleted from pSW11.	This work
pSW14	As pSW13 but with the ends blunted before religation.	This work
pSW15	pSW12 with an internal *Stu*I deletion.	This work
pTEP4	4.2kb *Eco*RV fragment from λ140 cloned into the same site in pBluescriptIISK	T. Pratt (this work)

2.2 GENERAL METHODS

2.2.1 General Procedures (Sambrook *et al.*, 1989)

All procedures were carried out at room temperature unless otherwise stated. All small-scale procedures were performed in 1.5ml polypropylene microcentrifuge tubes. Large scale procedures were performed in 15ml or 30ml Corex tubes or in 50ml Falcon tubes. Liquids were dispensed using Gilson P-20, P-200, and P-1000 air displacement pipettors and tips, or glass pipettes.

Microfuge tubes, pipette tips and glass pipettes were sterilised by autoclaving. Corex tubes were sterilised by baking at 250°C for 16 hours. Falcon tubes were supplied presterilised by the manufacturers.

2.2.2 Preparation of Dialysis tubing

Dialysis tubing (Visking 8/32") was prepared by boiling in a large volume of 2% (w/v) sodium bicarbonate, 1mM EDTA for 10 minutes and then in dH_2O for another 10 minutes. Dialysis tubing prepared in this way was stored in 50% ethanol at 4°C and rinsed in sterile dH_2O before use.

2.2.3 Phenol extraction of nucleic acid solutions

Proteins were removed from nucleic acid solutions by extraction with phenol. An equal volume of TE or citrate buffer-saturated phenol was added to the nucleic acid solution and the phases mixed by vortexing. The resulting mixture was then separated into organic and aqueous phases, by centrifugation at 11000g in a microfuge or Sorvall SS34 rotor for 3-5 minutes, and the upper (aqueous) phase transferred to a fresh tube. The process was repeated until the interface between the phases was free from denatured proteins. Phenol extraction was usually followed by one extraction with an equal volume of freshly prepared phenol:chloroform:isoamyl

alcohol (25:24:1), then with chloroform. The chloroform extraction step removes traces of phenol from the DNA solution, which might otherwise inhibit enzymes. Nucleic acids were then precipitated with ethanol (section 2.2.4).

2.2.4 Ethanol precipitation of nucleic acids

Ethanol precipitation of nucleic acids removes residual salts, phenol, and chloroform from nucleic acid solutions and is used to concentrate nucleic acids.

0.1 volumes of sodium acetate (pH 5.2) and 2-3 volumes of ethanol were added to the nucleic acid solution. The mixture was then incubated at -70°C for at least 15 minutes, and the precipitated nucleic acids were recovered by centrifugation at 11000g for 15 minutes at 4°C. The supernatant was discarded and the pellet was washed with 70% ethanol (v/v) to remove traces of the original solution, then recovered by centrifugation as before. The supernatant was drained off into paper and the pellet dried under vacuum. The pellet was then dissolved in the required volume of the appropriate buffer.

2.2.5 Quantitation of nucleic acid solutions

The concentration and purity of nucleic acid solutions were estimated by UV spectrophotometry at 260nm and 280nm. Concentration was estimated by assuming that an A_{260} of 1 corresponds to 50μg/ml for dsDNA and 40μg/ml for ssDNA and RNA. The ratio of A_{260}/A_{280} of a relatively pure solution of nucleic acids should be close to 2.

2.2.6 Autoradiography

Radiolabelled nucleic acids and proteins were detected by autoradiography. X-Ray film (DuPont Cronex or Amersham Hyperfilm) was exposed to gels or to nylon membranes containing [^{32}P]-labelled nucleic acids in lightproof cassettes by one of the following methods (in order of increasing sensitivity) :

i) at room temperature

ii) with calcium tungstate intensifying screens at -70°C

iii) as in ii) but with preflashed X-Ray film

Gels containing [^{32}P]-labelled run-off transcripts were exposed to an Imaging Plate (BAS-III) and analysed with a Bioimage analyser (BAS-2000, Fuji Film Co.).
[^{35}S]-labelled nucleic acids and proteins in fixed and dried gels were autoradiographed at room temperature.

2.3 BACTERIAL TECHNIQUES

2.3.1 Growth of bacterial strains

Except where otherwise stated, bacterial strains were grown in LB or LB agar containing the appropriate antibiotic for maintenance of plasmid and strain characteristics. Liquid cultures were routinely started by inoculating 5ml of LB plus antibiotic(s) with a fresh single colony. The M13 hosts NM570 and TG1 were plated on minimal medium when single colonies were required, in order to maintain F plasmid derivatives required for M13 sensitivity. Both strains carry a mutation affecting proline biosynthesis which is complemented by genes on the F plasmid derivative. For long term storage, a single colony was inoculated into 2ml of freezing broth containing the appropriate antibiotic(s) and grown overnight at 37°C.

1ml of sterile glycerol (50% w/v) was then added and the culture frozen at -70°C. Such preparations are stable at -70°C indefinately.

2.3.2 Transformation and transfection with DNA (Sambrook *et al.*, 1989)

Plasmid DNA and M13 replicative form DNA were introduced into *E.coli* strains by the same $CaCl_2$ based method. 50ml prewarmed LB was inoculated with 1/50 volume of a fresh overnight culture and shaken at 37°C until an A_{650} of 0.5-0.6 was reached. The culture was chilled on ice for 20 minutes, then split between two 30ml polycarbonate tubes and centrifuged (8000g) for 10 minutes at 4°C. The supernatant was discarded and pelleted cells resuspended in ice cold 0.1M $MgCl_2$ (10ml per tube) then centrifuged as before. The cells were then resuspended in ice cold $CaCl_2$ (10ml per tube), incubated on ice for 20 minutes, then centrifuged as before. The cells were then resuspended in 0.1M $CaCl_2$, 15% (v/v) glycerol (1ml per tube) and incubated on ice for 60 minutes. 0.2ml aliquots were transferred to microfuge tubes and stored at -70°C.

Transformation and transfection were effected by adding DNA (50-200ng plasmid or M13 RF DNA, or one quarter of a ligation mix) to 0.2ml of competent cells, freshly thawed, on ice. After 30 minutes incubation on ice, cells were heat-shocked (5 minutes at 37°C or 2 minutes at 42°C), then transferred back to ice for 10 minutes. Cells transformed by plasmid DNA were mixed with 1ml of LB and incubated at 37°C for 60 minutes prior to plating 0.2ml aliqouts on selective medium, in order to allow expression of antibiotic resistance.

When transfecting with M13 DNA, transfected cells were mixed with 0.2ml of a fresh overnight culture of the same strain, and then plated as for M13 titration (section 2.4.6).

2.4 PHAGE TECHNIQUES

2.4.1 Lambda titrations

20ml of prewarmed LB, supplemented with maltose (0.2% w/v), were inoculated with 1/20 volume of a fresh overnight culture of host bacteria and shaken vigorously at 37°C until an A_{650} of 0.5 was reached. Cells were harvested by centrifugation (2500g) in an MSE 'Mistral 1000' bench-top centrifuge and then resuspended in 10ml $MgSO_4$ (10mM) and stored at 4°C. Such "plating cells" can be used for up to one week.

Lambda lysates were serially diluted ten fold in phage buffer and 0.1ml aliquots of the appropriate dilutions mixed with 0.2ml of plating cells. The samples were then incubated at 37°C for 15 minutes to allow adsorption of the phages to the bacterial cells, then mixed with 3ml of BBL top agar and poured onto fresh, well dried BBL agar plates. The top agar was allowed to set, then the plates were incubated inverted at 37°C overnight.

2.4.2 Lambda plate lysates

A single fresh plaque was picked into 1ml of phage buffer using a sterile tooth pick. A small drop of $CHCl_3$ was then added and the phage suspension was incubated at 37°C for 15 minutes. 0.1ml of phage suspension and 0.2ml of plating cells were mixed and incubated at 37°C for 15 minutes. 3ml of molten BBL top agar, supplemented with maltose (0.2%) and $MgSO_4$ (10mM) were then added and the mixture poured onto a fresh, moist LB agar plate (also supplemented with maltose and $MgSO_4$).

The plate was incubated upright at 37°C until confluent lysis had occured (6-8 hours). The plate was then flooded with 4ml of LB and stored at 4°C overnight to allow phages to diffuse into the broth. The LB was then drawn off with a pipette,

mixed with a small drop of $CHCl_3$ and clarified by centrifugation (8000g) in a Sorvall SS34 rotor at 4°C for 10 minutes.

2.4.3 Lambda liquid lysates

200ml of prewarmed LB, supplemented with maltose (0.2% w/v) and $MgSO_4$ (10mM) were inoculated with 1/40 volume of a fresh overnight host culture and shaken vigorously at 37°C. When an A_{650} of 0.5 was reached, phage lysate was added to give a multiplicity of infection of 0.5, and incubation was continued until lysis occurred. 0.5ml of chloroform was added and the culture incubated with shaking for 10 more minutes at 37°C. 8g of NaCl were then dissolved in the culture, and RNase A and DNase I added to a final concentration of 1μg/ml each. After one hour of incubation at room temperature the phage suspension was clarified by centrifugation (8000g) for 20 minutes at 4°C. The supernatant containing the phages was then decanted carefully into a conical flask and stored at 4°C . At this stage the lysate was titrated.

Phages were concentrated by precipitation with PEG, as follows. PEG was dissolved in the phage lysate to a final concentration of 10% (w/v) by slow mixing on a magnetic stirrer at room temperature, then the mixture was transferred to ice for 60 minutes to allow precipitation of phages and macromolecules. Precipitated phage particles were then recovered by centrifugation (9000g) for 20 minutes at 4°C. The supernatant was discarded and the phage pellet gently resuspended in 5ml of phage buffer. PEG and cell debris were removed from the phage suspension by extraction with 1 volume of chloroform. The organic and aqueous phases were then separated by centrifugation (3000g) in a Sorvall SS34 rotor at 4°C for 15 minutes and the aqueous phase transferred to a fresh tube.

2.4.4 Lambda phage purification

Lambda phage particles were purified by equilibrium centrifugation in caesium chloride. After extraction with $CHCl_3$ the volume of the phage suspension was measured and CsCl added (0.75g per ml of lysate). When the CsCl had dissolved completely the phage suspension was transferred to an 11ml crimp-seal centrifuge tube and centrifuged at 90,000g for 24 hours at 4°C.

The phage particles were visible as an opaque band in the CsCl solution and were recovered through the side of the tube using a hypodermic syringe fitted with a 19g needle. CsCl was then removed from the phage suspension by dialysis against two changes of phage buffer for two hours at room temperature. The purified phage suspension was then stored at 4°C in tightly capped microfuge tubes.

2.4.5 Lambda transduction

Mutations in cloned DNA were moved from plasmid vectors to the *E.coli* chromosome using lambda clones of the Kohara library (Kohara *et al.,* 1987) as specialised transducing phages. This method was devised and described in full by Kulakauskas *et al.* (1991).

The location of the cloned gene on the physical map of the *E.coli* K12 chromosome must be known in order that the appropriate lambda clone can be used in the procedure. The plasmid borne copy of the gene is then mutated by insertion of an antibiotic resistance marker between appropriate restriction site(s). Lambda phages are then grown on cells carrying the mutated plasmid and the resulting lysate used for transduction. Phages of the Kohara library are *cI*, therefore transductants recovered are haploid recombinants produced by recombination between homologous sequences in the cloned and chromosomal alleles.

A plate lysate of the appropriate Kohara phage was prepared by infecting TG1 cells, carrying mutant plasmids, as described in section 2.4.2 then titrated before storage at 4°C.

An overnight culture of the recipient strain was then grown and subcultured as for lambda titrations. When an A_{650} of 0.5 was reached the culture was infected with phage lysate at an m.o.i. of 1 and incubated with shaking 37°C for 15 minutes. Trisodium citrate and glucose were then added to a final concentration of 10mM and 0.2% respectively. The culture was then grown at 37°C for 2 hours to allow expression of antibiotic resistance. 0.1ml aliquots were then mixed with 3ml of BBL top agar and plated on LB agar containing appropriate antibiotic(s), then incubated at 37°C overnight.

Transductants were screened for the absence of plasmids by loss of antibiotic resistance and by small scale plasmid DNA preparation. Candidate transductants were then purified and characterised.

2.4.6 M13 titration

Phages were ten fold serially diluted in phage buffer. 10µl of each phage dilution were mixed with 0.2ml of a fresh overnight culture of TG1 and added to 3ml of molten BBL top agar, containing IPTG (166µg/ml) and X-gal (250µg/ml) for blue/white selection. The mixture was poured onto a fresh, dry BBL plate and allowed to set. Plates were incubated inverted at 37°C overnight.

2.4.7 Storage of M13 phages

Single plaques were toothpicked into 0.5ml aliquots of phage buffer and stored for several months at 4°C. Alternatively, aliquots of supernatant from phage infected cultures were stored at 4°C. Fresh, single plaques were obtained by plating dilutions of phage stock as described in section 2.4.6. For long term storage, phage infected cells were frozen at -70°C as described in section 2.3.1.

2.5 NUCLEIC ACID PURIFICATION

2.5.1 Small scale plasmid DNA purification (Birnboim and Doly, 1979)

Plasmid DNA was isolated from small volume cultures by alkaline lysis, which selectively denatures and precipitates chromosomal DNA and proteins. The following solutions were used in this method :

Solution I	50mM glucose, 25mM Tris-HCl pH 8.0, 10mM EDTA pH 8.0.
Solution II	0.2M NaOH, 1% (w/v) SDS freshly prepared, as required.
Solution III	3M potassium acetate, 2M glacial acetic acid.

1.5ml of an overnight culture (previously grown from a single colony, in LB plus appropriate antibiotic) were transferred to a microfuge tube and centrifuged (11000g) for 2 minutes at room temperature. The supernatant was discarded and the pellet resuspended in 100µl of solution I (to which 4mg/ml lysozyme had been added) and incubated at room temperature for 5 minutes. 200µl of solution II were added and the sample mixed by inversion, then incubated on ice for 5 minutes. 150µl of solution III were added and the sample left on ice for a further 5 minutes to allow precipitation of SDS, chromosomal DNA, and proteins. The precipitate was then pelleted by centrifugation (11000g) for 5 minutes and the supernatant transferred to a fresh tube. The sample was extracted with an equal volume of phenol:chloroform:isoamylalcohol. Nucleic acids were concentrated by ethanol precipitation and vacuum dried. The dried pellet was then redissolved in 30µl TE containing 20µg/ml RNase A.

2.5.2 Large scale isolation of plasmid DNA

200ml of LB, plus appropriate anitibiotic, were inoculated with a single colony and shaken at 37°C overnight. Cells were harvested by centrifugation (8000g) in a Sorvall GSA rotor at 4°C for 20 minutes. The cells were then resuspended in 100ml of TE and pelleted as before. The pellet was resuspended in 3ml of sucrose buffer (25% (w/v) sucrose, 40mM Tris-HCl pH 8.0) and placed on ice. 1ml of lysozyme (10mg/ml) was added and the sample incubated on ice for 5 minutes. 1µl of 0.5M EDTA (pH 8.0) and 800µl of RNase A (5mg/ml) were added and the sample was incubated on ice for a further 5 minutes. 5ml of Triton mix (0.1% (w/v) Triton X-100, 65mM EDTA pH 8.5, 50mM Tris-HCl pH 8.0) were added and the sample incubated for 10 minutes on ice. Cell debris was pelleted by centrifugation (27000g) in a Sorvall SS34 rotor for 30 minutes. The supernatant was transferred to a fresh tube and its volume adjusted to 9ml. 9g of CsCl were added and allowed to dissolve, after which 0.9ml of EtdBr (5mg/ml) were added. The sample was then transferred to a crimp-seal polyallomer tube and centrifuged (90,000g) in a Beckman Ti50 rotor for 60 hours at 18°C. The DNA was observed by fluorescence of the DNA-EtdBr complex under UV light. Two bands were visible. The lower band, almost exclusively comprising supercoiled plasmid DNA, was recovered through the side of the tube using a hypodermic syringe and needle. EtdBr was removed by four extractions, each with an equal volume of propan-2-ol saturated CsCl solution. The DNA solution was dialysed against several changes of TE buffer for a total of 24 hours at 4°C. The DNA was finally concentrated by ethanol precipitation, vacuum dried, and redissolved in 100µl TE buffer.

2.5.3 Purification of M13 Replicative Form (RF) dsDNA

Small scale

A fresh single plaque was toothpicked into 1.5ml of LB and shaken at 37°C overnight. The culture was then transferred to a 1.5ml microfuge tube and the cells harvested by centrifugation. RF DNA was prepared from infected cells by the same method used for small scale plasmid purification.

Large scale

200ml of LB were inoculated with 1/50 volume of an overnight liquid culture (grown from a single colony of the host strain) and shaken at 37°C until an A_{650} of 0.45 was reached. 0.2ml of supernatant from a 1.5ml phage culture was then added and the culture grown for a further 4 hours. RF DNA was purified from phage infected cells by the same method used for large scale plasmid purification.

2.5.4 Purification of M13 ssDNA

Small scale

Phages were grown as for small scale RF preparation. The culture was then transferred to a microfuge tube, and bacterial cells pelleted by centrifugation (11000g) for 5 minutes. 1ml of supernatant was transferred to a fresh tube and 0.2ml of 20% (w/v) PEG 8000, 2.5M NaCl added. The tube was vortexed and placed at 4°C for 15 minutes, then precipitated phages were pelleted by centrifugation (11000g) for 5 minutes. The supernatant was discarded, and remaining traces of supernatant removed with a rolled up tissue. The phage pellet was resuspended in 0.1ml of TE. 0.1ml of phenol was added and the tube vortexed several times over 5

minutes, then centrifuged for 5 minutes. 80µl of the aqueous phase were transferred to a fresh tube and ssDNA recovered by ethanol precipitation.

Large scale

This method was taken from the site directed mutagenesis kit instruction booklet (Amersham International plc). The phage inoculum was prepared by toothpicking a single plaque into 1.5ml of LB and shaking overnight at 37°C. Bacterial cells were removed by centrifugation and the supernatant stored at 4°C. 100ml of LB were inoculated with 1ml of host strain culture (grown overnight from a single colony) and shaken at 37°C until an A_{650} of 0.3 was reached. 1ml of phage inoculum was then added and the culture grown for a further 4 hours. Cells were then pelleted by centrifugation (8000g) in a Sorvall GSA rotor at 4°C for 30 minutes. The supernatant was carefully transferred to a fresh tube and 20ml of 20% (w/v) PEG, 2.5M NaCl were added. The sample was mixed and stored at 4°C for at least 1 hour, then centrifuged as before. The supernatant was discarded, then the sample was centrifuged for 5 more minutes and remaining traces of supernatant removed with a drawn out pasteur pipette. The pellet containing the phages was resuspended in 0.5ml of TE and transferred to a microfuge tube, then centrifuged (11000g) for 5 minutes to pellet any remaining bacterial cells. The supernatant was transferred to a centrifuge tube, mixed with 0.2ml of 20% (w/v) PEG, 2.5M NaCl, and incubated at room temperature for 15 minutes. Precipitated phages were harvested by centrifugation (11000g) for 5 minutes and the supernatant discarded. The phage pellet was resuspended in 0.5ml TE, mixed with 0.2ml of phenol by vortexing, then left to stand for 15 minutes at room temperture. The tube was then vortex mixed once more and centrifuged (11000g) for 5 minutes. The aqueous phase was transferred to a fresh tube and re-extracted with phenol. The sample was then extracted once with phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1). ssDNA was then recovered by ethanol precipitation and the dried pellet redissolved in 50µl TE.

2.5.5 Purification of Bacteriophage Lambda DNA (Sambrook *et al.*, 1989)

DNA was purified from concentrated phage lysates, prepared as in section 2.4.4, by phenol extraction. 0.5ml aliquots of phage lysate were mixed by several inversions with an equal volume of phenol, then left to stand at room temperature for 15 minutes. Each sample was vortexed again and centrifuged (11000g) for 5 minutes. The aqueous phase was transferred to a fresh tube and phenol extracted once more, then once with phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1). The aqueous phases were then pooled and dialysed against several changes of TE buffer at 4°C overnight. At this stage the DNA concentration was measured by UV spectrophotometry at 260nm. If necessary the DNA was further concentrated by precipitation with ethanol.

2.5.6 Purification of *E.coli* genomic DNA

This protocol is modified from Hermann and Frischauf (1987).

100ml of LB (plus appropriate antibiotic) were inoculated with 1/50 volume of an overnight culture and shaken at 37°C until an A_{650} of 0.5 was reached. Cells were harvested by centrifugation (8000g) in a Sorvall GSA rotor at 4°C for 10 minutes and the supernatant discarded. Cells were then resuspended in 20 ml of STE (25% sucrose, 50mM Tris-HCl pH 8.0, 10mM EDTA pH 8.0) and transferred to a Falcon tube. 1ml of SDS (10% w/v) and 0.4ml proteinase K (10mg/ml) were added and the tube was transferred to a 50°C waterbath for 6 hours. The sample was then mixed with 5ml of phenol by gently inverting several times and left to stand at room temperature for 15 minutes. The tube was then centrifuged for 20 minutes at 4°C and the aqueous phase transferred to a fresh tube. The sample was then gently extracted twice with phenol:chloroform:isoamyl alcohol and the DNA recovered by precipitation with ethanol. The DNA pellet was dried by standing the tube at room temperature until the last traces of ethanol had evaporated. The DNA was then resuspended in 1ml of TE and stored at 4°C until dissolved.

2.5.7 Purification of RNA from *E.coli* cells

This protocol is taken from Salser *et al.* (1967) and Shaw and Guest (1982).

200ml of Spizizen minimal medium were inoculated with 1/50 volume of an overnight culture (grown from a single colony) and shaken at 37°C until an A_{650} of 0.3 was reached.

Alternatively, cells were grown to various stages of stationary phase (early, mid, and late) as determined by the A_{650} of diluted samples. Cells were cold-shocked by transferring to a 10°C waterbath and samples were taken every hour for 4 hours, then after 24 hours. For osmotic-shock induction, the culture was treated with NaCl (at 0.5M final concentration) and grown for a further 30-60 minutes (Yim *et al.*, 1994). Cells were harvested by centrifugation (8000g) for 10 minutes at 4°C. The supernatant was discarded and the cells resuspended in 20ml of TE, then centrifuged as before. The pellet was resuspended in 1ml of ice-cold 10mM KCl, 4mM $MgCl_2$, 10mM Tris-HCl pH 7.3, and transferred to a microfuge tube. Lysozyme was then added to a final concentration of 300µg/ml and the sample frozen at -70°C for 30 minutes. Upon thawing, 110µl of 10% (w/v) SDS were added and the tube incubated at 64°C until the turbidity of the solution had dropped (usually 5-10 minutes). 40µl of 3M sodium acetate (pH 5.2) was added and the sample split between 2 microfuge tubes. Each sample was mixed with one volume of phenol and placed at 64°C for 5 minutes. During this incubation period the samples were mixed several times. The tubes were then centrifuged (11000g) for 5 minutes and the aqueous phase transferred to a fresh tube. The samples were phenol extracted once more as before, then precipitated with ethanol. The RNA pellet was washed 4 times in 70% (v/v) ethanol, 10mM Tris-HCl pH 7.5, 10mM NaCl.

2.6 MANIPULATION OF DNA

Unless otherwise stated, methods described in the following section are modifications of methods described in Sambrook *et al.* (1989).

2.6.1 Restriction endonuclease digestion

DNA was digested using a 2-10 fold excess of restriction endonuclease in a volume of 20-50µl for 1-3 hours. For restriction mapping purposes, reaction conditions were those recommended by the manufacturer. For general cloning purposes and where digestion with two enzymes (for which recommended reaction conditions differed) was required, reactions were usually performed in 1 x Universal buffer (33mM Tris-acetate pH 7.9, 10mM Mg-acetate, 66mM K-acetate, 5mM DTT, 0.1mg/ml BSA). Where appropriate, enzymes were inactivated by heating at 75°C for 15 minutes, or by extraction with phenol followed by ethanol precipitation.

2.6.2 Dephosphorylation of DNA

Vector DNA used in cloning experiments was dephosphorylated after digestion with restriction enzymes in order to minimise the recovery of parental clones. 0.5-3µg vector DNA, in a final volume of 20µl, were digested with the required restriction enzyme(s) as described in section 2.6.1. Thirty minutes before the end of the reaction 0.1 unit of calf intestinal alkaline phosphatase was added and the reaction completed at 37°C. The reaction was then heated at 75°C for 10 minutes to inactivate the phosphatase and nucleases, then extracted with phenol. DNA was recovered by precipitation with ethanol.

2.6.3 Electrophoresis of DNA

For restriction analysis, Southern blotting, and restriction fragment purification, DNA molecules were size-separated by electrophoresis in submerged agarose gels. All gel formers, combs, and electrophoresis tanks were bought from Bethesda Research Laboratories (Gaithesburgh, MD, USA).

Gels were 0.6-3% (w/v) agarose in 1 x TAE or 1 x TBE, with the same buffer in the electrophoresis tank. For rapid separation, gels were cast in small (50mm x 75mm) gel formers and run with an electric field of 2-5 V/cm. For accurate mapping and restriction fragment purification, gels were cast in larger (11 x 14 cm) gel formers and run with an electric field of 0.75-1.25 V/cm. Prior to loading, restriction digests were heated at 75°C for 15 minutes to inactivate enzymes and separate cohesive ends, then mixed with 0.1 volumes of agarose gel loading buffer (see section 2.1.2). After electrophoresis, gels were stained in EtdBr (1μg/ml) for 30 minutes, destained in dH_2O for the same time and viewed on a UV transilluminator. Photographs were taken with a Polaroid Land Camera fitted with a red filter, or with a UVP electronic camera connected to a Sony Digital Graphic Printer.

DNA molecules were also size-separated by electrophoresis in nondenaturing acrylamide gels for purifying small (< 500bp) restriction fragments. Gels were 4-8% acrylamide in 1 x TBE with 1 x TBE buffer in the electrophoresis tank. Electrophoresis tanks, combs and gel formers were obtained from the ATTO corporation, Tokyo, Japan. The following solutions were used in this method :

Acrylamide stock solution	30% (w/v) acrylamide, 0.8% N,N'-methylene bis-acrylamide, filtered through Whatman No. 1 paper and stored in the dark at 4°C.
AMPS	10% (w/v) ammonium sulphate, freshly made.

To prepare a gel, the appropriate volume of stock acrylamide solution was mixed with 3ml 10 x TBE, 50µl AMPS and dH_2O to 30ml final volume. 20µl TEMED were added and the mixture poured into a 16 x 16 x 0.1cm mould formed by two glass plates, one notched at one end, the other square, separated by 0.1cm spacers and sealed round the edges with a rubber gasket. The mould was filled to the top and a 14 tooth comb inserted into the top of the gel. Polymerisation was allowed to proceed for 30 minutes, after which time the comb and gasket were removed. The gel was attached to the electrophoresis tank fitted with 1 x TBE running buffer and the wells were washed out. DNA samples were heated, mixed with 10 x TBE loading buffer and loaded onto the gel. Samples were run at 1-8 V/cm. After electrophoresis the plates were separated and the gels were stained with 0.5µg/ml EtdBr for 30 minutes. DNA bands were visualised on a UV transilluminator.

2.6.4 Purification of restriction fragments from agarose gels

Restriction fragments of DNA used in cloning experiments were purified from TAE agarose gels using the Gene Clean kit bought from Bio 101 Inc. (La Jolla, CA, USA).

Briefly : The DNA was digested with the appropriate enzyme(s) and the products separated by electrophoresis in a TAE agarose gel (section 2.6.3). The EtdBr stained gel was examined on a UV trasilluminator and a gel slice containing the required restriction fragment excised using a scalpel. The agarose was dissolved in 2.5 volumes of 6M sodium iodide at 55°C, then mixed with the appropriate volume of "glass milk" suspension (usually 5µl) and incubated at room temperature for 5 minutes to allow adsorption of the DNA to the silica matrix of the glass milk. The silica matrix and bound DNA were pelleted by centrifugation then washed 3 times in "New Wash" (a solution of Tris/NaCl/EDTA in ethanol). The DNA was eluted from the silica matrix by washing it twice at 55°C in a volume of TE equal to that of the glass milk.

DNA prepared by this method can be used directly in cloning procedures.

2.6.5 Purification of restriction fragments from polyacrylamide gels

The following procedure is a modification of the "Crush and Soak" method described by Sambrook *et al.* (1989).

Restriction fragments were run on a nondenaturing polyacrylamide gel as described in section 2.6.3 and the EtdBr-stained bands visualised on a UV transilluminator. The required fragment was cut from the gel using a scalpel, chopped into small pieces, and transferred to a microfuge tube. The gel fragments were crushed against the wall of the tube using a disposable pipette tip and 2 volumes of elution buffer (0.5M ammonium acetate, 10mM Mg-acetate, 1mM EDTA pH 8.0, 0.1%SDS) were added. The tube was closed, sealed with parafilm, and incubated at 37°C overnight on a rotary wheel. The sample was then centrifuged (11000g) for 2 minutes and the supernatant transferred to a fresh microfuge tube. An additional 0.5 volumes of elution buffer were added to the pellet and the sample vortexed and centrifuged as above. The supernatants were combined and extracted twice with phenol and once with phenol:chloroform:isoamyl alcohol (25:24:1). The DNA was ethanol precipitated at -70°C for 30 minutes and recovered by centrifugation. The pellet was redissolved in 200µl TE, reprecipitated with ethanol, and washed in 70% ethanol. The DNA was dried at room temperature and resuspended in 10µl TE.

2.6.6 Ligation of DNA

Ligation reactions were routinely performed in a final volume of 15µl containing 200-300ng of dephosphorylated cut vector DNA, a five-fold excess of "insert" DNA, 1.5µl of 10 x ligation buffer (0.4M Tris-HCl pH 7.6, 0.1M $MgCl_2$, 0.1M DTT, 500µg/ml BSA, 10mM ATP), DNA ligase and dH_2O.

For ligation of cohesive ends , 0.5-2 units of T4 DNA ligase were used and the reaction allowed to proceed for 4 hours at room temperature or 16 hours at 16°C. Blunt end ligations were performed using 5 units of T4 ligase for 16 hours at 16°C.

2.7 DNA SEQUENCING

DNA sequence was determined by modifications of the chain termination method (Sanger *et al.,* 1977) using the "Sequenase" T7 DNA polymerase sequencing kit (U.S.B.) according to the manufacturer's instructions. A synthetic oligonucleotide primes synthesis of a new strand of DNA in two stages. A labelling reaction is performed in the presence of limiting concentrations of all four dNTPs, one of which is radiolabelled. This produces a population of radio-labelled DNA molecules ranging from a few nucleotides to several hundred nucleotides in length. In the second step, four parallel reactions are performed, one for each of the four bases (dA, dC, dG, dT). DNA synthesis is continued in each reaction, in the presence of higher concentrations of all four dNTPs plus one ddNTP. Provided the dNTP:ddNTP ratio is correct, a fraction of the population of extending DNA chains is terminated at each site where the ddNTP can be incorporated.

The products of all four reactions are size separated in adjacent tracks on denaturing polyacrylamide gels and visualised by autoradiography.

2.7.1 Sequencing of ssDNA

Annealing primer to template

1-3μg of ss M13 DNA and 0.5-1pmol of oligonucleotide primer were mixed with 2μl of 5 x Sequenase reaction buffer (200mM Tris-HCl pH 7.5, 100mM $MgCl_2$, 250mM NaCl) and dH_2O in a final volume of 10μl. The mixture was incubated at 65°C for two minutes then allowed to cool to room temperature for 15-30 minutes. It was then transferred to ice.

Labelling reaction

The annealed template-primer was mixed with 1µl of DTT (0.1M), 0.5µl of [α-^{35}S] dATP (10µCi/µl) and 2µl of labelling mix (1.5µM each of dCTP, dGTP, and dTTP). 2µl of sequenase T7 polymerase (diluted 1/8 in 10mM Tris-HCl pH 7.5, 5mM DTT, 0.5µg/ml BSA) were added and the mixture incubated for 5 minutes at room temperature.

Termination reactions

Four microfuge tubes were labelled A, C, G, and T. 2.5µl of the appropriate termination mix (containing 80µM of each dNTP plus 8µM of one ddNTP) were added to each tube and the tubes transferred to a 37°C waterbath to prewarm.
When the labelling reaction was complete, 3.5µl aliquots of labelling reaction were mixed with each of the termination mixes and incubated at 37°C for 5 minutes. The reactions were stopped by addition of 5µl of formamide dye (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). The tubes were then transferred to a 75°C waterbath for 2-5 minutes and loaded immediately onto a denaturing polyacrylamide gel (see section 2.7.3).

2.7.2 Sequencing of dsDNA

Double-stranded plasmid or M13 RF DNA can be sequenced directly provided the template is first denatured to allow annealing of the primer.
3-5µg of plasmid DNA were denatured in 1M NaOH (made up freshly from 10M stock) and annealed with the primer in a final volume of 10µl for 15 minutes at 37°C. The solution was neutralised by addition of 1M HCl, and sequencing was then performed as for single stranded templates.

2.7.3 Sequencing gels

The products of sequencing reactions and primer extensions were size-separated by electrophoresis through Tris-borate-urea sequencing gels. All gels were run using the components of the Base Runner sequencing apparatus manufactured by IBI, using the method supplied in the manufacturer's instruction booklet. Products of run-off transcription were also analysed on denaturing polyacrylamide gels using the same apparatus as for nondenaturing gels (section 2.6.3).
The following solutions were used in this method :

1 x TBE 6% gel mix	150ml 40% acrylamide stock, 100ml 10 x TBE, 460g urea, dH_2O to 1L.
40% acrylamide stock	38% (w/v) acrylamide, 2% (w/v) N,N'-methylene bis acrylamide deionised and filtered through Whatman No. 1 filter paper.
AMPS	10% (w/v) ammonium persulphate.

Gels were prepared by mixing 50ml 1 x TBE 6% gel mix with 150µl AMPS and 50µl TEMED and poured, using a 25ml glass pipette, into a 60cm x 21cm x 0.4mm mould formed by two glass plates, one notched, one rectangular and insulated, separated down the long edges by 1cm wide, 0.4mm thick spacers and sealed with special "underwater" pvc tape round the edges. The flat side of a sharkstooth comb (32 lanes) was inserted 0.5cm into the top of the gel which was clamped at the corners and allowed to set for 30 minutes.

When polymerisation was complete the tape and comb were removed, the gel sandwich was mounted on the electrophoresis apparatus and the upper and lower buffer tanks were filled with 1 x TBE. The top of the gel was then washed thoroughly with 1 x TBE using a syringe and needle, and the comb was re-inserted with its teeth just piercing the surface of the gel. The gel was pre-run at 45W constant power for 30 minutes, then the wells were washed out with 1 x TBE before

the samples were loaded. The gel was then run at 45W constant power for between 2 and 6 hours depending on the range of sequence to be read.

After the run, the plates were separated and the gel was fixed in 10% (v/v) methanol, 10% (v/v) glacial acetic acid for 20 minutes. The gel was then transferred to a sheet of blotting paper, dried at 80°C using a heated vacuum drier and autoradiographed.

2.8 NUCLEIC ACID HYBRIDISATION

2.8.1 3′ end-labelling of DNA fragments with Klenow

Recessed 3′ termini of double-stranded DNA molecules can be filled using the Klenow fragment of *E.coli* DNA polymerase I. This is a modification of the protocol described by Sambrook *et al.* (1989).

DNA restriction fragments were purified from agarose or nondenaturing acrylamide gels as described in sections 2.6.4 and 2.6.5. 0.5-1.0μg were added to 2μl of 10 x Universal buffer, 1μl of a mixture of dATP, dGTP, and dTTP (2mM each), 1μl [α-^{32}P] dCTP (10mCi/ml, 3000Ci/mmol), 1 unit of Klenow and dH_2O to 20μl final volume. After 15 minutes at room temperature the reaction was chased with 1μl of unlabelled 2mM dCTP and incubated for a further 5 minutes. The reaction was stopped by heating for 5 minutes at 70°C and then transferred to ice. Labelled DNA was separated from unincorporated dNTPs by two rounds of ethanol precipitation followed by a 70% ethanol wash.

2.8.2 5′ end-labelling of oligonucleotide probes

20-100pmol oligonucleotide were mixed with 2.5μl 10 x kinase buffer (0.5M Tris-HCl pH 8.0, 0.1M $MgCl_2$), 2μl DTT (0.1M), 0.5-2.5μl [γ-^{32}P] ATP (10mCi/ml, 3000Ci/mmol), 20 units T4 polynucleotide kinase and dH_2O to 25μl. The mixture

was incubated at 37°C for 30 minutes, then the enzyme was inactivated by heating at 75°C for 15 minutes. The labelled oligonucleotide was stored at -20°C.

2.8.3 Transfer of DNA to nylon membranes (Southern blotting)

This protocol is taken from the Hybond N^+ Instruction Booklet, Amersham International plc.

Digested DNA was electrophoretically separated in agarose gels as described in section 2.6.3. The gel was stained and photographed, then immersed in denaturing buffer (0.5M NaOH, 1.5M NaCl) and incubated at room temperature, with shaking, for 30 minutes. The gel was rinsed briefly in dH_2O, then immersed in two changes of neutralising buffer (1.5M NaCl, 0.5M Tris-HCl pH 7.2, 1mM EDTA) for 15 minutes each, and finally rinsed in dH_2O.

To prepare the blotting platform a glass plate, supported by an small inverted plastic tray, was placed in the middle of a shallow glass dish, filled with blotting buffer (20 x SSC or 20 x SSPE). A wick was formed by placing three sheets of Whatman 3MM paper (soaked in blotting buffer) on top of the glass plate and in contact with the buffer reservoir.

The gel was inverted onto the wick and a piece of Hybond N^+ positively charged nylon membrane, cut to the correct size, was placed on top of the gel, followed by three sheets of Whatman 3MM paper soaked in blotting buffer and a stack of absorbent paper towels. The gel was then surrounded with Saran Wrap to prevent blotting buffer being absorbed directly by the paper towels. A glass plate was placed on top of the paper stack and weighed down with a small (0.75 kg) weight. Transfer of DNA from the gel to the nylon membrane was allowed to proceed for 2-16 hours. DNA was then immobilised on the membrane by alkali fixation :

The membrane was placed on a pad of Whatman 3MM paper saturated with NaOH (0.4M) for 30 minutes, then rinsed briefly in 5 x SSC or 5 x SSPE and allowed to dry at room temperature. The membrane was then wrapped in Saran Wrap and stored at 4°C.

Alternatively the blotting buffer was replaced with 0.4M NaOH and DNA transfer performed as above. After 2-16 hours the membrane was rinsed, dried and stored as above without the need for further fixation.

DNA samples were dot blotted onto nylon membranes by heating the sample to 95°C then chilling on ice. 1 volume of 20 x SSC was added and samples spotted onto the membrane in 2µl aliquots. The membrane was placed, DNA side up, on a pad of Whatman 3MM paper soaked in denaturing buffer (as above) for 5 minutes, then on a pad of Whatman 3MM paper soaked in neutralising buffer (as above) for 1 minute. The DNA was fixed to the membrane as described above.

2.8.4 Hybridisation of DNA probes to DNA immobilised on nylon membranes

Membranes prepared as in section 2.8.3 were prehybridised with 10ml of SSC hybridisation buffer (6 x SSC, 20% v/v formamide, 5 x Denhardt's solution, 0.5% SDS, 50 mM sodium pyrophosphate, 0.5mg/ml heparin from porcine intestinal mucosa) in a Techne hybridisation cylinder at 65°C for 1 hour. Prehybridisation is necessary in order to block non-specific sites on the DNA and nylon membrane, to which the probe and free radiolabelled nucleotides would otherwise bind. This would result in a high background signal which would obscure the signal from any sequence-specific binding by the probe.

After prehybridisation was completed the solution was discarded and fresh hybridisation buffer containing denatured radiolabelled probe was added to the cylinder. Denaturation of the probe was achieved by boiling for 5 minutes immediately before use. Hybridisation was allowed to proceed at 65°C for 12-18 hours. The hybridisation solution was then removed and stored at -20°C for re-use. The membrane was washed briefly at room temperature in 6 x SSC, then twice at 65°C, for 20 minutes each, in 1 x SSC, 0.1% w/v SDS. The membrane was then wrapped in Saran Wrap and autoradiographed for 12-24 hours. More stringent washes (higher wash temperature and/or lower salt concentration) or longer exposure times were performed where necessary.

2.8.5 Hybridisation of oligonucleotide probes to immobilised DNA

Oligonucleotide probes were hybridised to the immobilised DNA at 10°C below the calculated melting temperature (T_M) of the DNA-oligonucleotide hybrids. The T_M was estimated using the equation of Meinkoth and Wahl (1984) for oligonucleotides up to 17bp:

$$T_M\,(°C) = 4(G + C) + 2(A + T)$$

where (G + C) = number of G:C base pairs and (A + T) = number of A:T base pairs.

Filters were prehybridised with 10ml SSC hybridisation buffer (as section 2.8.4) for 1 hour at 10°C below the calculated T_M. The hybridisation solution was discarded and fresh solution containing the probe was added. Hybridisation was allowed to proceed for 18-24 hours. The hybridisation solution was then removed and stored at -20°C. The filter was rinsed briefly in 6 x SSC at room temperature then washed three times in 6 x SSC, 0.1% w/v SDS, for 15 minutes each, at 30°C. Finally, the filter was washed twice in 6 x SSC, 0.1% w/v SDS at T_M -10°C for five minutes each, then wrapped in Saran Wrap and autoradiographed. More stringent washes or longer exposure times were performed when necessary.

2.8.6 Removal of probes from nylon membranes

DNA probes and oligonucleotides can be removed from a nylon membrane that has been kept moist, thereby allowing re-use of the membrane in further hybridisations. The membrane was immersed in boiling 0.5% SDS (w/v) and allowed to cool slowly to room temperature. The membrane was then autoradiographed to check that the probe had been removed successfully. If clean it was then allowed to dry at room temperature, then wrapped in Saran Wrap and stored at 4°C. Membranes kept in this way were prehybridised and hybridised with new probes as described (section 2.8.5).

2.8.7 Transfer of RNA to nylon membranes (Northern blotting)

This protocol is taken from the Hybond N^+ Instruction Booklet, Amersham International plc.

RNA was purified from *E.coli* cells as described in section 2.5.7 and incubated at 65°C for 5 minutes in the following solution : 6µl RNA, 12.5µl formamide (deionised), 2.5µl 10 x MOPS buffer (0.2M 3-[N-Morpholino]-propane-sulphonic acid, 0.05M Na acetate pH 7.0, 0.01M EDTA), 4µl 37% formaldehyde. The mixture was chilled on ice then 2.5µl of 50% (v/v) glycerol containing 0.1mg/ml bromophenol blue was added. This was then run on a 1-1.5% denaturing agarose gel made up in 1 x MOPS buffer and 13% formaldehyde, with 1 x MOPS as running buffer. The RNA was then transferred to a Hybond N^+ membrane by 20 x SSC capillary blotting as described in section 2.8.3. Alkali fixation was achieved by placing the membrane on a pad of Whatman 3MM filter paper soaked in 0.05M NaOH for 5 minutes. The membrane was rinsed briefly in 2 x SSC and then hybridised with oligonucleotide probes as described in section 2.8.5, or wrapped in Saran Wrap and stored at 4°C.

RNA samples were dot blotted onto nylon membranes by incubating at 65°C in three volumes of the following solution : 500µl formamide, 162µl formaldehyde (37% solution), 100µl 10 x MOPS buffer. After chilling on ice, one volume of 20 x SSC was added and samples spotted onto the membrane in 2µl aliquots. RNA was fixed to the membrane as described above, then hybridised with oligonucleotide probes as described in section 2.8.5.

2.8.8 Primer extension analysis of RNA

This protocol is a modification of the method described by Sambrook *et al.* (1989).

Primer extension was used to map the 5′ termini of mRNA molecules. The test RNA was hybridised with an excess of a 5′ end-labelled oligonucleotide primer, then the primer extended by reverse transcription. The length of the cDNA product(s), as

judged by denaturing polyacrylamide gel electrophoresis, was a measure of the distance between the 5′ labelled nucleotide of the primer and the 5′ terminus of the mRNA.

20-100μg RNA prepared as described in section 2.5.7 were mixed with 0.5-5pmol oligonucleotide primer (end-labelled as described in section 2.8.2). 3μl formamide (special biochemical grade, BDH), 6μl 5 x First Strand Buffer (250mM Tris-HCl pH 8.3, 375mM KCl, 15mM $MgCl_2$, Gibco BRL) and dH_2O to 30μl final volume were added and the mixture incubated at 85°C for 10 minutes in order to denature nucleic acids. The mixture was then incubated at 5-10°C below the estimated T_M of the oligonucleotide (see section 2.8.5) for 60 minutes to allow hybridisation to occur. The sample was then chilled on ice and mixed with 4μl 5 x First Strand Buffer, 5μl DTT (0.1M), 2.5μl dNTP mix (10mM each of dATP, dCTP, dGTP, dTTP), 1μl actinomycin D (1mg/ml, Sigma), 20 units RNase inhibitor (Boehringer), 200 units M-MLV RTase (Mouse Moloney Leukemia Virus Reverse Transcriptase, Gibco BRL), and dH_2O to 50μl final volume. The mixture was then incubated at 37°C for 60 minutes. After this time, 12.5μl NaOH (0.5M) were added and the sample boiled for 3 minutes to hydrolyse RNA. 12.5μl each of HCl (0.5M) and 1M Tris-HCl (pH 7.4) were added and the sample precipitated with ethanol. The dried DNA pellet was dissolved in 4μl dH_2O and mixed with 6μl formamide dye (see section 2.7.1). Samples were then boiled for 3 minutes and the cDNA products size-separated on DNA sequencing gels (see section 2.7.4). A DNA sequencing ladder, generated using the same primer, was run along side as a marker.

2.8.9 Nuclease S1 analysis of RNA

Nuclease S1 was used to map the 3′ termini of mRNA on ds DNA templates. Double-stranded DNA is denatured under conditions that minimise the formation of DNA : DNA hybrids while promoting the formation of DNA : RNA hybrids. Under these conditions, the coding region within a DNA fragment hybridises with its corresponding mRNA while the remainder of the DNA remains single-stranded.

DNA that has not formed duplexes is hydrolysed by nuclease-S1, whereas DNA that has hybridised to RNA is protected from digestion. Double-stranded DNA fragments that are 3′-end labelled within the gene of interest can be size-separated on a denaturing acrylamide gel following RNA hybridisation and nuclease-S1 digestion to map the 3′ of the mRNA. This protocol is modified from that described by Sambrook *et al.* (1989).

Double-stranded DNA fragments carrying a recessed 3′ end within the coding region of the gene of interest were purified from nondenaturing acrylamide gels as described in section 2.6.5 and 3′ end-radiolabelled as described in section 2.8.1. 0.1-1.0μg DNA and 10-250μg RNA (prepared as described in section 2.5.7) were mixed, ethanol precipitated, and dried. The pellet was resuspended in 30μl hybridisation buffer (40mM PIPES [piperazine-N,N′-bis(2-ethanesulphonic acid)] pH 6.4, 1mM EDTA pH 8.0, 0.4M NaCl, 60% formamide) and the mixture incubated at 85°C to denature the nucleic acids. This was then rapidly transferred to a waterbath set at 52-55°C (as determined for a particular DNA : RNA hybrid by a series of preliminary reactions) and hybridisation allowed to continue for 12-16 hours. After this time 300μl of ice-cold nuclease-S1 mapping buffer (0.28M NaCl, 0.05M Na-acetate pH 4.5, 4.5mM $ZnSO_4$, 20μg/ml single-stranded DNA, 100-1000 units/ml nuclease-S1) were added and the mixture immediately vortexed. This was then transferred to a 37°C waterbath and incubated for 1-3 hours depending on the degree of digestion required. The reaction was chilled on ice and stopped by the addition of 80μl stop mixture (4M ammonium acetate, 50mM EDTA pH 8.0, 50μg/ml tRNA) followed by vortexing. The reaction was extracted once with phenol : chloroform, centrifuged, and the aqueous phase was ethanol precipitated. The dried pellet was resuspended in 10μl TE and mixed with 10μl formamide dye (section 2.7.1). After 2-5 minutes in a 75°C waterbath 5μl were run on a denaturing polyacrylamide gel as described in section 2.7.3.

2.9 *IN VITRO* TRANSCRIPTION

These assays were performed at the National Institute of Genetics, Mishima, Japan, under the supervision of Professor Akira Ishihama.

Promoter selectivity by reconstituted RNA polymerase holoenzymes $E\sigma^{70}$ and $E\sigma^{S}$ were determined *in vitro* by single round run-off transcription. DNA fragments carrying the promoter region of interest were purified from nondenaturing polyacrylamide gels as described in section 2.6.5. 0.1pmol of promoter DNA (as determined by UV spectrophotometry) was mixed on ice with 3.5µl 10 x transcription buffer (500mM Tris-HCl, 500mM NaCl, 30mM Mg-acetate, 1mM EDTA, 1mM DTT, 250µg/ml BSA), 1pmol of core RNA polymerase, and 4pmol of either σ^{70} or σ^{S} in a final volume of 35µl. The mixture was preincubated at 37°C for 15-30 minutes to allow open complex formation. 15µl of prewarmed substrate-heparin mixture (1.5µl 10 x transcription buffer, 2µl 25 x XTP mixture [4mM ATP, GTP, CTP, plus 1.25mM UTP], 2µCi [α- ^{32}P] UTP, 2µl 5mg/ml heparin, dH_2O to 15µl) were then added (without removing the preincubation mixture from the waterbath) and incubated for a further 5 minutes. 50µl of ice-cold stop solution (40mM EDTA, 300µg/ml tRNA) were added and the sample immediately vortexed. Nucleic acids were precipitated by adding 500µl ethanol and 100µl 4M ammonium acetate and incubating the sample at -70°C for 30 minutes. After centrifugation (11000g) for 10 minutes the pellet was washed in 70% ethanol and resuspended in 15µl formamide dye (see section 2.7.1). This was incubated at 90°C for 2 minutes then placed on ice. The whole sample was loaded onto a 6% denaturing polyacrylamide gel as described in section 2.7.3 and labelled RNA products were analysed with a Bio-Imaging Analyser.

2.10EXPRESSION AND LABELLING OF PLASMID-ENCODED PROTEINS

2.10.1 Expression of proteins in minicells

E.coli strains which carry mutations in the *min* locus divide asymmetrically, resulting in the formation of anucleate cells (minicells) which are capable of supporting DNA synthesis, transcription, and translation, but which contain no chromosomal DNA. Plasmids segregate efficiently into minicells; therefore purified minicells can be used to analyse proteins under essentially *in vivo* conditions.

Purification of minicells

Three 500ml batches of LB plus appropriate antibiotic(s) were inoculated with 1ml of a fresh overnight culture of the minicell-producing strain P678.54 transformed with the required plasmid, and shaken overnight at 37°C. At the same time eight sucrose gradients were prepared by freezing 25ml aliquots of 20% w/v sucrose in M9 minimal medium in 30ml glass corex tubes at 70°C, then thawing them undisturbed overnight at 4°C.

The next day, most viable cells were removed from the culture by centrifugation (1000g) in a Sorvall GS3 rotor for 15 minutes at 4°C. Minicells were then harvested from the supernatant by centrifugation (12000g) in a Sorvall GS3 rotor for 20 minutes at 4°C. The pellets were resuspended in a total volume of 10ml of supernatant and transferred to a 30ml corex tube. The tube was then vortexed vigorously three times, for one minute each, to disperse aggregated minicells which would otherwise pellet with remaining viable cells in subsequent centrifugation steps. 2.5ml aliquots of minicell suspension were layered on top of each of four 35ml sucrose gradients and centrifuged (4000g) in a Sorvall HB-4 swing-out rotor for 20 minutes at 4°C. The minicells formed a diffuse band within the sucrose gradient while the viable cells were pelleted. The top two-thirds of each minicell

band (approximately 10ml) were removed and pooled samples from pairs of gradients pelleted by centrifugation in a Sorvall SS34 rotor (16000g) for 10 minutes at 4°C. The pellet was resuspended in 4ml of M9 medium and purified twice more in sucrose gradients as before (using two gradient tubes each time). The final minicell pellet was resuspended in 1ml 30% (w/v) glycerol in M9 medium and the A_{600} of the sample measured. The sample was then aliquoted into 1.5ml microfuge tubes such that the volume of each when diluted to 1ml gave an A_{600} of 0.2 (2×10^9 minicells). Aliquots were stored at -70°C.

Labelling of plasmid-encoded proteins

Minicell aliquots were thawed at room temperature and diluted to 1ml final volume in M9 glucose medium. The minicells were pelleted by centrifugation then resuspended in 0.1ml of M9 glucose medium and incubated at 37°C for 60 minutes to ensure degradation of any mRNA still present in the minicells. 20μCi of [^{35}S]-methionine (10μCi/μl, 1000 Ci/mmol) in Difco methionine assay medium (25% w/v in M9 glucose) were added and the samples incubated at 37°C for 3 hours (or at 10°C for 1-4 hours for cold-shock experiments). Samples were then chased with 5μl unlabelled methionine (8mg/ml) for 3 minutes, and the minicells were pelleted by centrifugation for 3 minutes. Minicells were then washed in 50mM Tris-HCl (pH 6.8), pelleted as before, and resuspended in the residual liquid. 15μl of cracking buffer (section 2.10.1) were added and the samples boiled for 3 minutes. Samples were then fractioned in SDS polyacrylamide gels as described in section 2.11. Labelled proteins were detected by autoradiography.

2.10.2 Induction of T7 RNA polymerase promoters

Male *E.coli* cells susceptible to M13 infection were transformed with plasmids containing the T7 promoter controlling the gene to be expressed. A single colony from a fresh minimal agar plate was grown overnight with antibiotic selection, diluted 1/100 in LB containing antibiotic and grown for several hours at 37°C with gentle shaking. At an A_{650} of 0.5, 3ml of cells were harvested by centrifugation and the cells resuspended in 5ml M9 minimal medium. The cells were harvested again and the pellet resuspended in M9 minimal medium plus 18 amino acids (0.01% each minus methionine and cysteine). The cells were then grown for 1-2 hours at 37°C with gentle shaking, after which time they were infected with M13 mGP1-2 at an m.o.i. of 10. mGP1-2 carries the T7 RNA polymerase gene under the control of the *lac* promoter. IPTG was added at the same time (to a final concentration of 1mM) which induces expression of the T7 RNA polymerase gene. After 30 minutes at 37°C rifampicin was added to a final concentration of 200μg/ml. After a further 30 minutes, 1ml of cells were pulsed with 10μCi of [^{35}S]-methionine for 5 minutes. Cells were harvested in a microfuge and resuspended in 120μl cracking buffer (60mM Tris-HCl pH 6.8, 1% β-mercaptoethanol, 10% glycerol, 3% SDS, 0.01% bromophenol blue). Samples were heated to 95°C for 3 minutes and a 40μl aliquot loaded onto an SDS polyacrylamide gel (see section 2.11).

2.10.3 Linked T7 transcription-translation system

This protocol is taken from the Linked T7 transcription-translation system Instruction Booklet, Amersham International plc.

This system is designed for efficient *in vitro* synthesis of proteins from a variety of templates containing T7 promoters upstream from coding sequences, including supercoiled plasmids, linearised plasmids, and PCR products. The system is based on transcription with T7 RNA polymerase followed by translation in an optimised rabbit reticulocyte lysate.

All plasmid DNA templates were purified on CsCl gradients as described in section 2.5.2. Linearised templates were purified from agarose gels as described in section 2.6.4. 0.5μg of template DNA were added to 8μl of T7 transcription mix and the volume made up to 10μl with nuclease-free water. After mixing gently by pipetting up and down, the transcription reaction was incubated for 15 minutes in a 30°C waterbath. 30μl of translation mix and 1-4μl of [^{35}S]-methionine (10μCi/μl, 1000 Ci/mmol) were then added, and the final volume made up to 50μl with nuclease-free water. After gentle mixing the translation reaction was incubated in a 30°C waterbath for 60 minutes. 50μl of cracking buffer (see section 2.10.1) were added and the samples boiled for 3 minutes. Samples were then fractioned in SDS polyacrylamide gels as described in section 2.11 and labelled proteins detected by autoradiography.

2.11 SDS POLYACRYLAMIDE GEL ELECTROPHORESIS

Protein samples were analysed by SDS polyacrylamide gel electrophoresis, using a discontinuous buffer system, as described by Laemmli (1970). The following solutions were used in this method :

Acrylamide stock solution	30% (w/v) acrylamide, 0.8% (w/v) N,N'-methylene bis-acrylamide filtered through Whatman No.1 paper and stored in the dark at 4°C.
Resolving gel buffer	1.5M Tris-HCl pH 8.8, 0.4% (w/v) SDS.
Stacking gel buffer	0.5M Tris-HCl pH 6.8, 0.4% (w/v) SDS.
AMPS	10% (w/v) ammonium persulphate, made fresh daily.
Sample loading buffer	50mM Tris-HCl pH 6.8, 1% (w/v) SDS, 10% (v/v) glycerol, 0.05% (w/v) bromophenol blue, 1% (v/v) β-mercaptoethanol.

To prepare a 10% gel : 10ml acrylamide stock solution , 7.5ml resolving gel buffer, 50µl AMPS and dH_2O to 30ml were mixed. 20µl TEMED were added and the resolving gel mixture poured into a 16 x 16 x 0.1cm mould formed by two glass plates, one notched at one end, the other one square, separated by 0.1cm spacers and sealed round the edges with a rubber gasket. The mould was fitted to within 5cm of the notched plate, then the gel solution was carefully overlayed with H_2O-saturated butan-2-ol. The gel was then allowed to polymerise for thirty minutes in an upright position.

To prepare the 10% stacking gel mix : 3ml acrylamide stock were mixed with 2.25ml stacking gel buffer, 25µl AMPS and dH_2O to 9ml.

The H_2O-saturated butan-2-ol was poured off the polymerised resolving gel and the surface washed with dH_2O. 10µl TEMED were then added to the stacking gel solution and it was poured into the gel mould until it reached the top of the notched plate. A 14 tooth comb was then inserted into the top of the gel solution and polymerisation was allowed to proceed for 30 minutes. The comb and rubber gasket were then removed and the gel attached to an electrophoresis tank fitted with 1 x Tris-glycine running buffer containing 0.5% (w/v) SDS.

The wells were washed out with running buffer and any air bubbles trapped beneath the gel were allowed to escape by gently tipping the whole apparatus to one side. Protein samples prepared as described in sections 2.10.1 to 2.10.3 were then loaded and the gel run at 60-90V overnight.

Gels of varying percentages were prepared in the same way by altering the proportions of acrylamide stock and dH_2O in both the stacking and resolving gel mixes.

After running, the gel was removed from the plates and fixed and stained for 30 minutes at 37°C in a solution of 50% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.1% Coomassie brilliant blue. The gel was then destained in 3-4 changes of 10% (v/v) glacial acetic acid over several hours at 37°C.

Once destained the gels were transferred to blotting paper and dried on a heated vacuum drier. Gels containing [^{35}S]-labelled proteins were then autoradiographed as described in section 2.2.6.

2.12 COMPUTER METHODS

The University of Wisconsin Genetics Computer Group (GCG) sequence analysis software package, mounted on a UNIX operating system, was used in most sequence analyses. The MAP and CODONFREQUENCY programs were used to obtain restriction maps and translated DNA sequences, and to aid analysis of codon choice displayed by open reading frames respectively. PEPTIDESORT was used to calculate protein molecular weights and amino acid composition, and PILEUP to perform multiple alignments of protein sequences.

BLAST, FASTA, and BESTFIT programs were used to find DNA and protein sequence similarities. The EMBL and GenBank sequence data libraries were used as a source of DNA sequences of interest whilst SWISS-PROT and NBRF-PIR provided relevant protein sequence information. BLAST and FASTA programs were also used to find *E.coli* sequence similarities within the ECDC database, WWW URL : http://susi.bio.uni-giessen.de/usr/local/www/html/ecdc.html.

Chapter 3

3.1 Introduction

The identification of alternative sigmas has clearly established their importance in the regulation of gene expression. Their ability to confer a new promoter specificity on RNAP provides an efficient means of co-ordinately activating unlinked genes in response to a particular environmental stimulus. As described in chapter 1 at least six alternative sigmas regulate the expression of various stress response genes in *E.coli*. Whether further sigmas control other regulons remains an important question in our understanding of bacterial gene expression.

The majority of the known *E.coli* alternative sigma factors were identified as positive regulators of gene expression in response to various stresses (section 1.8). In an attempt to identify new sigmas, Fujita *et al.* (1987) utilised the available data on sigma sequence homology provided by Gribskov and Burgess (1986) to design a synthetic tetradecameric peptide matching exactly the amino acid sequence of the highly conserved subregion 2.2 (section 1.5.4). The authors raised polyclonal antiserum against the synthetic peptide and screened for proteins carrying antigenic epitopes similar to subregion 2.2. The anti-peptide antiserum was found to cross-react specifically with both σ^{70} and σ^{32} on Western blots, and with about 10 other *E.coli* proteins termed sigma cross-reacting proteins or SCRPs (Fujita *et al.*, 1987; Ueshima *et al.*, 1992). Four major SCRPs were purified and N-terminally sequenced and two of them, SCRP-27B and SCRP-34, were identified as the ribosomal protein S2 and thioredoxin reductase respectively. SCRP-23 was later identified as the C22 component of alkyl hydroperoxide reductase encoded by *ahpC* (Smillie *et al.*, 1992) although the function of SCRP-27A remains unknown. None of these SCRPs was found to bind core RNAP; none was found exclusively in the RNAP holoenzyme fraction upon glycerol gradient fractionation of crude lysates of *E.coli* proteins, and none showed competition with σ^{70} for immunoprecipitation (Fujita *et al.*, 1987; Ueshima *et al.*, 1992). In addition, none share significant amino

acid homology with sigma proteins (including subregion 2.2). Thus it is virtually certain that they do not function as sigma factors.

As another approach to identifying candidate sigma genes, work in this laboratory has used oligonucleotides corresponding to conserved regions within sigma factors to probe the library of Kohara *et al.* (1987), which consists of physically mapped *E.coli* genome fragments in λ vectors. In this chapter I describe the mapping, sequencing and characterisation of a candidate sigma gene, *f229*, which was identified by this approach.

3.2.1 Design of oligonucleotide probes

Two 78-mer oligonucleotide probes corresponding to the conserved core-binding subregion 2.1 of bacterial sigma factors (Lonetto *et al.,* 1992) were designed in this laboratory by David A. Smillie. The degeneracy of the genetic code means that the target DNA sequence cannot be exactly defined from the corresponding amino acid sequence. However, the stablity of hybrids formed by long oligonucleotide probes can outweigh the destabilizing effects of mismatches within the probe/target complementarity (Lathe, 1985). Since most amino acids are specified by codons that differ only at the third position, at least two out of three nucleotides of such codons are guaranteed to be a perfect match. Where possible, regions lacking leucine, arginine and serine are chosen since each of these amino acids is specified by six codons.

The two probes were designed on the basis of codon choice as described by Ikemura (1985). One probe, 746J, carried the least optimal codon for each amino acid within the subregion 2.1 target sequence, while the other, 745J, carried the codons recognised by major tRNA species of *E.coli.* The amino acid of σ^{32} was chosen for positions of little or no conservation. It was thought that additional alternative sigmas may well be encoded by poorly translated mRNA (see table 3.1)

Table 3.1. F_{op} values of *E.coli* sigma genes.

The codon usage of each gene was analysed using the Codonfrequency program of the GCG package. The frequency of optimal codons (F_{op}) was calculated by dividing the number of optimal codons by the total number of optimal plus non-optimal codons assigned according to Ikemura (1985).

Gene	F_{op}
rpoD	0.72
rpoS	0.68
rpoH	0.81
rpoE	0.60
rpoF	0.61
fecI	0.51
rpoN	0.62

An F_{op} value of 0.60 to 0.62 shows that there is no strong bias in the coding sequence of these genes towards codons recognised by major tRNA species of *E.coli*. Such values are characteristic of moderately to weakly expressed *E.coli* genes. The high F_{op} for *rpoH* may be an additional feature of its complex regulation which results in enhanced synthesis of σ^{32} following temperature upshift (see section 1.8.3). The low F_{op} for *fecI* may reflect the fact that its product is responsible for the recognition of only one promoter (see section 1.8.5), and so high levels of translation would be wasteful.
It should be noted that *rpoE* was unsequenced and FecI was not known to be a sigma factor when the probes described in section 3.2.1 were designed.

so as not to significantly reduce levels of $E\sigma^{70}$, hence the use of both major and minor codon probes. As such, the oligonucleotides differed at 16 positions out of a total of 78 nucleotides. The conserved amino acid sequence of subregion 2.1 of alternative bacterial sigmas (corresponding to, although not identical to residues 371 to 396 of *E.coli* σ^{70}) is : ADQEARKTLIEANLRLVVSIAKKYAN (as determined from Lonetto *et al.,* 1992).

Based on the minor codons of this sequence, the nucleotide sequence of 746J is :

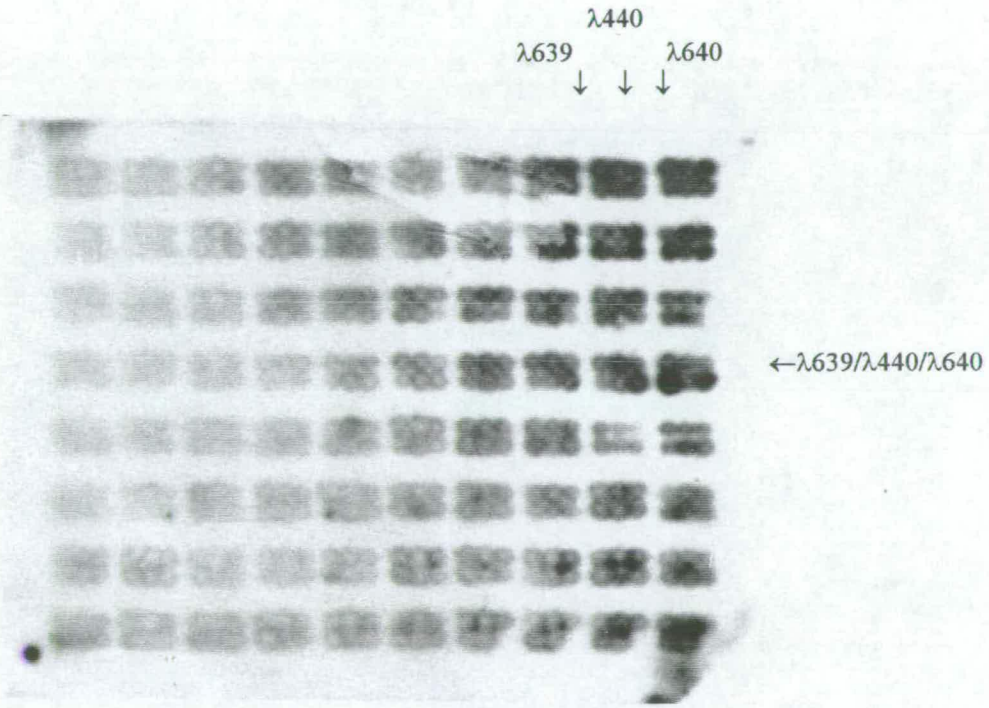
5'-GCC GAT CAG GAA GCC CGC AAA ACC CTG ATT GAA GCC AAT CTG CGC CTG GTG GTG TCG ATT GCC AAA AAA TAT GCC AAT -3'

3.2.2 Probing *E.coli* genomic DNA

746J was 5' end-labelled using [γ-^{32}P] ATP as described in section 2.8.2 and hybridised with nylon filters carrying DNA from the miniset clones of the *E.coli* K12 genome (Kohara *et al.,* 1987; Noda *et al.,* 1991). This probe strongly hybridised with λ440 and more weakly with a pair of contiguous clones, λ639 and λ640 (fig. 3.1). The 19kb *E.coli* insert DNA of λ440 is centred near 2775kb on the physical map (59.4 minutes on the genetic map). The inserts of λ639 and λ640, which overlap by approximately 9kb, are centred near 4322kb (92.5 minutes). Although a 2.8kb probe-positive *Pst*I fragment was identified in a λ440 DNA digest (data not shown), this and other positive fragments were unclonable (D.Smillie, personal communication). λ440 is discussed in section 3.10. No other mapped sigma genes hybridised with this probe.

The location of a candidate sigma gene in the overlap region of the λ639/λ640 insert DNA was investigated further.

Figure 3.1. Location of candidate sigma genes identified using oligonucleotide 746J within the Kohara miniset clones of the *E.coli* genome. Hybridisation (section 2.8) was performed at 50°C for 18 hours. Final washing conditions were 2xSSC, 0.1% SDS for two periods of 5 minutes at 50°C. The positions of λ clones 440, 639 and 640 are marked. λ639 is present at the bottom right position of the panel indicated. λ440 and λ640 are present within the adjacent panel; λ440 is at the bottom left position and λ640 is at the bottom right position. (Layout of the Kohara miniset library is such that each rectangular "panel" contains six phages, all with the same final two digits in their identifying numbers, but different initial digits. (This ensures that neighbours originate from far apart on the chromosome). Numbering runs left to right, and then top to bottom. Thus for example the tenth rectangle in the fourth row contains phages 140, 240 *et seq*).



3.2.3 Fine mapping of the λ639/λ640 probe positive region

The probe-binding region was mapped more accurately by Southern analysis of λ639 and λ640 restriction fragments as described in section 2.8.3. DNA was prepared from purified λ639 and λ640 lysates (previously amplified on the *recD* strain NM621, which is particularly suitable for the propagation of the *spi*⁻ recombinant phages of the Kohara library) as described in section 2.5.5. Restriction analysis using the "Kohara enzymes" confirmed the identity of λ639 and λ640 (fig. 3.2) and gave results consistent with the published map of the *E.coli* K12 chromosome (Kohara *et al.,* 1987). As determined from fig. 3.3, 746J hybridised to a region within the overlap of λ639 and λ640 flanked by a *Bam*HI site and a *Pvu*II site. The *Bam*HI site at 4322.6kb is in the *gltP* gene encoding a proton-glutamate-aspartate transport protein (Wallace *et al.,* 1992), while the *Pvu*II site at 4324.2kb was found to lie in the intergenic region between *gltP* and *fdhF*, the latter encoding a selenocysteine-containing subunit of formate dehydrogenase (Zinoni *et al.,* 1986). The λ639/λ640 insert region is shown in fig. 3.4. The absence of significant sequence homology between 746J and the published sequence of *gltP* (Tolner *et al.,* 1992), as determined by the Fasta and Bestfit programs of the GCG package, indicated that the probe positive region lies in the ~1.3kb unsequenced *gltP/fdhF* intergenic region.

3.3 Subcloning the probe positive region

A 3.7kb *Bam*HI-*Kpn*I fragment from λ639 carrying the 3′ ends of *gltP* and *fdhF* was subcloned between the corresponding sites in pBluescript KS(-) to form pSW3. The structure of pSW3 was verified by restriction analysis. In addition, a spontaneous deletion lacking ~200bp of insert DNA was obtained and called pSW4 (fig. 3.5). For sequencing purposes, the same *Bam*HI-*Kpn*I fragment was subcloned into the polylinkers of M13 tg130 and M13 tg131 to give tg130-4 and tg131-3.

Figure 3.2. Restriction analysis of λ639 and λ640 DNA on a 0.6% agarose gel. Tracks 2,3,5,7,9, and 11 carry λ639 DNA; tracks 4,6,8,10,12, and 13 carry λ640 DNA. The following digests were loaded: track 2, *Eco*RI; tracks 3 and 4, *Eco*RV; tracks 5 and 6, BglI; tracks 7 and 8, *Kpn*I ; tracks 9 and 10, *Pvu*II; track 11, *Kpn*I/*Pvu*II; track 12, *Bam*HI; track 13, *Pvu*II/*Bgl*I. Tracks 1 and 14 carry λcI857 *S*am7 marker DNA digested with *Hin*dIII and *Eco*RI/*Hin*dIII respectively. The relevant marker bands are indicated and their sizes given in kilobase pairs.

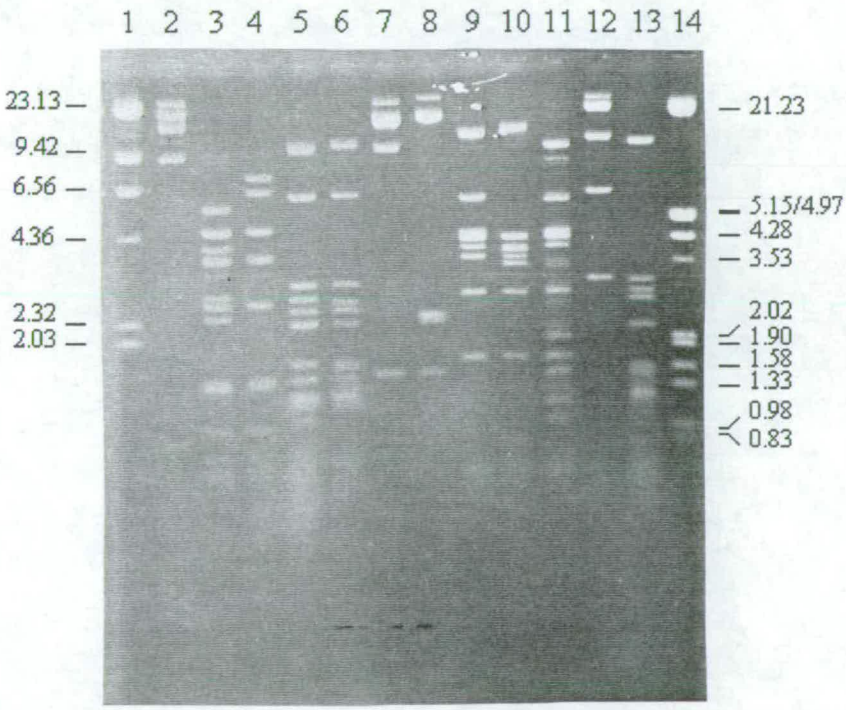


Figure 3.3. Southern analysis of λ639 and λ640 DNA using [γ-^{32}P]-labelled 746J as probe. DNA was transferred from the gel shown in fig. 3.2 to a nylon membrane as described in section 2.8.3. Hybridisation was performed at 50°C for 18 hours. Final washing conditions were 0.2xSSC, 0.1% SDS for 2 minutes at 65°C.

Arrows indicate probe-positive *Pvu*II (tracks 9,10,11, and 13) and *Bam*HI (track 12) fragments (2.8kb and 6.7kb respectively) which define the limits of the smallest positively-hybridising region. The sizes of the *Hin*dIII marker fragments are given at the left of the figure.

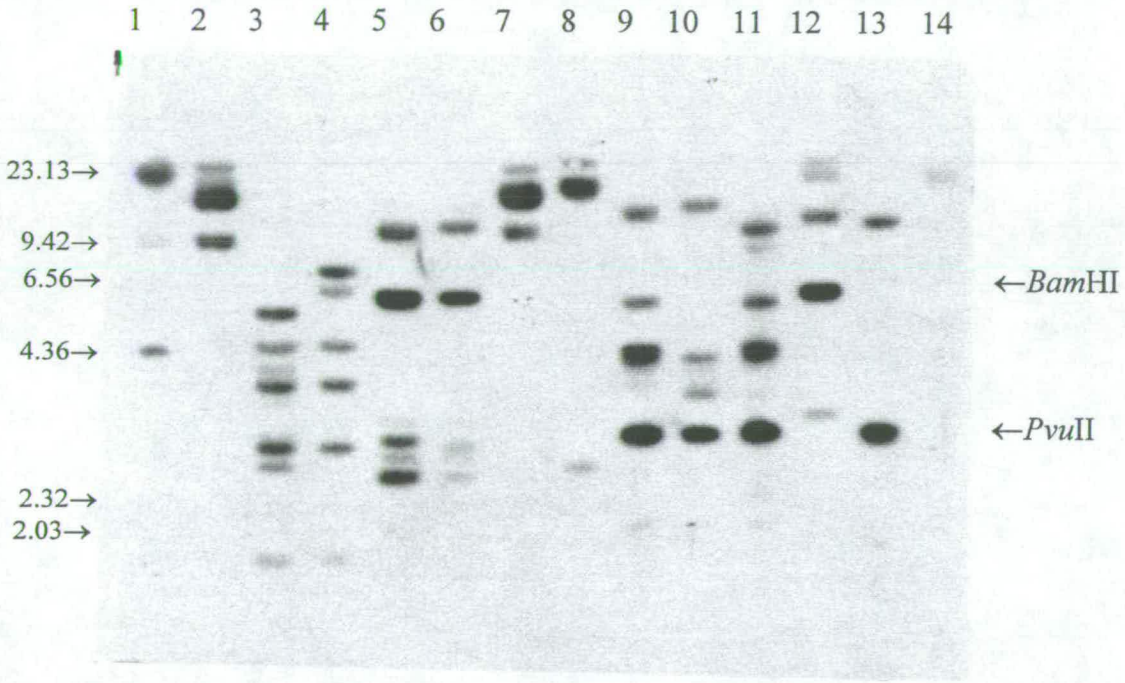


Figure 3.4. Restriction map of *E.coli* chromosomal inserts cloned in λ639 and λ640, based on the map of Rudd (1991). The top scale shows the coordinates in kb on the physical map of the *E.coli* K12 chromosome; the bottom scale shows the coordinates in minutes on the genetic map. The locations of *gltP* and *fdhF* are indicated. *Kpn*I and *Bam*HI restriction sites which flank the probe-positive region and were used to generate pSW3 are marked with asterisks, and the length and orientation (relative to the vector DNA) of the chromosomal inserts carried by λ639 and λ640 are indicated by arrows below the map. The following abbreviations are used for restriction enzymes: B, *Bam*HI; E, *Eco*RV; H, *Hin*dIII; G, *Bgl*II; K, *Kpn*I; P, *Pst*I; R,EcoRI; V, *Pvu*II.

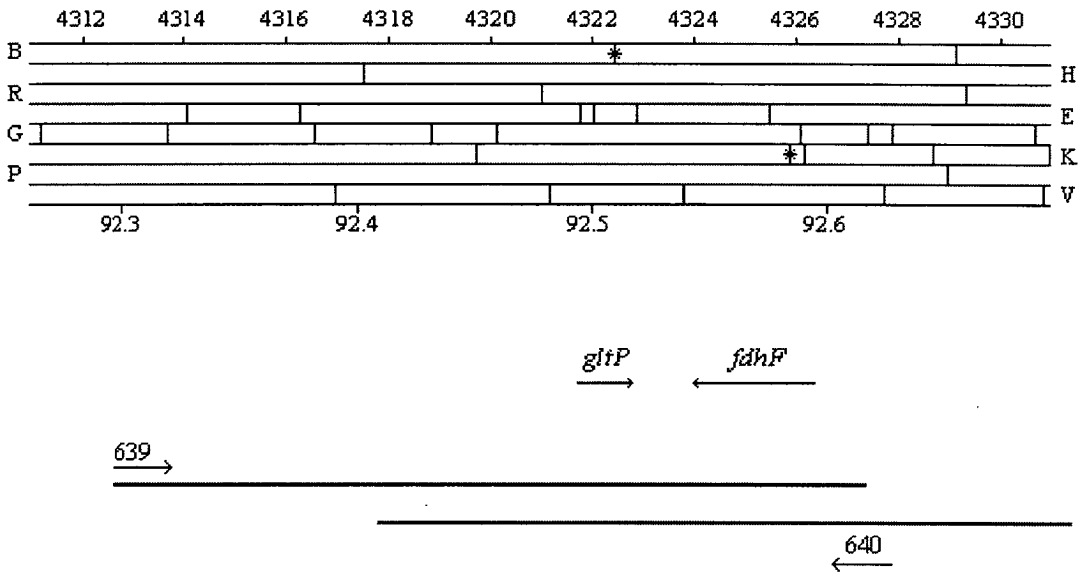
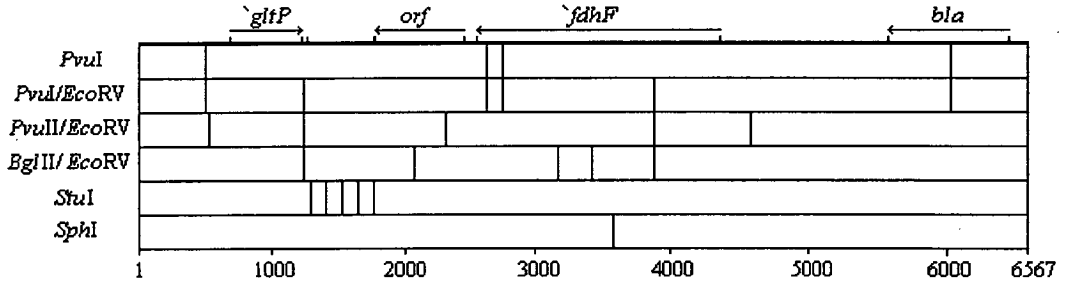
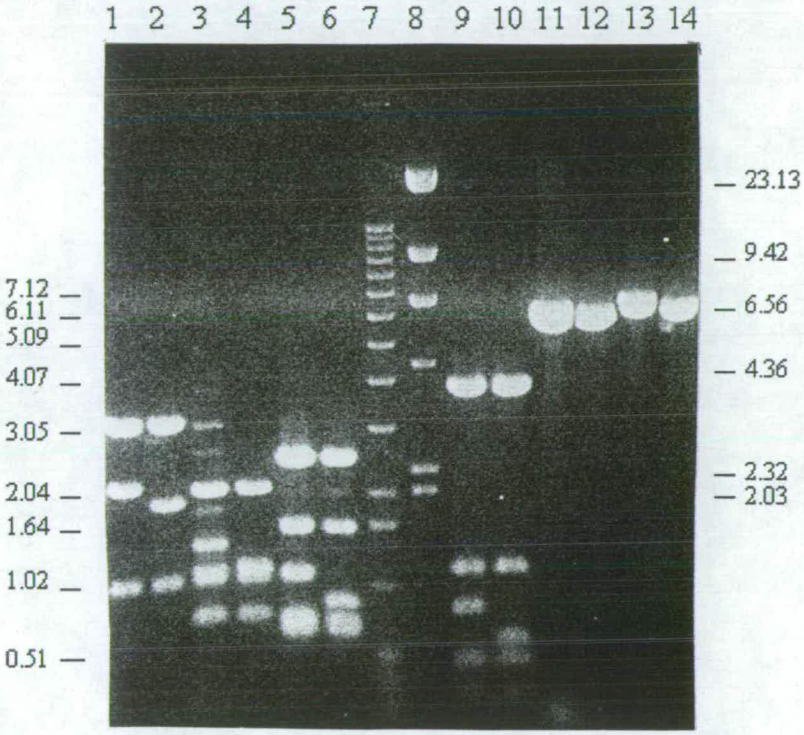


Figure 3.5. Restriction fragments of pSW3 and pSW4 DNA analysed by 0.6% agarose gel electrophoresis. Tracks 1,3,5,9,11, and 13 contain pSW3 DNA; Tracks 2,4,6,10,12, and 14 contain pSW4 DNA. Track 7 contains 1kb marker DNA (Gibco BRL); Track 8 contains λ*cI*857 *S*am7 marker DNA digested with *Hin*dIII. Relevant fragments are marked and their sizes given (in kilobase pairs) at the sides.

Restriction enzymes used are as follows : Tracks 1 and 2, *Pvu*I; Tracks 3 and 4, *Pvu*I+*Eco*RV; Tracks 5 and 6, *Pvu*II+*Eco*RV; Tracks 9 and 10, *Bgl*II+*Eco*RV; Tracks 11 and 12, *Stu*I; Tracks 13 and 14, *Sph*I.

Below is the corresponding restriction map of pSW3, linearised at position 1 of pBluescript KS(-). The bottom scale shows the distance in base pairs




1 2 3 4 5 6 7 8 9 10 11 12 13 14
23.13
9.42
6.56
4.36
2.32
2.03
7.12
6.11
5.09
4.07
3.05
2.04
1.64
1.02
0.51

3.4 Nucleotide sequence of the *gltP/fdhF* intergenic region

Single-stranded M13 DNA was prepared from tg130-4 and tg131-3 as described in section 2.5.4. Nucleotide sequencing was initiated using synthetic oligonucleotide primers corresponding to the 3′ ends of the published sequences of *gltP* and *fdhF*, as described in section 2.7. Further DNA sequence information was then generated using primers corresponding to already determined nucleotide sequences.

Sequencing downstream from within *gltP* using tg130-4 DNA consistently gave multiple pile-ups (fig. 3.6) which could not be overcome by addition of formamide to prevent non-specific primer hybridisation, by cycle sequencing at high temperature to alleviate template structural problems, or by using a different oligonucleotide primer (data not shown). Sequencing downstream from within *fdhF* using tg131-3 DNA gave no such problems. Forward sequences were confirmed by sequencing back towards *fdhF* using tg130-4 DNA showing that DNA template preparations were not responsible for the pile-ups in fig. 3.6. The sequencing strategy is shown in fig. 3.7, and the nucleotide sequence in fig. 3.8.

An open reading frame (ORF) predicted to be transcribed in the same orientation as *fdhF* was identified. This would extend for 229 amino acids if translation were initiated at the first AUG codon, although no convincing Shine-Dalgarno ribosome-binding sequences were found to precede this. A GA-GT sequence having 4 out of 7 bp of the consensus Shine-Dalgarno sequence AGGAGGT was found to precede a potential TTG initiation codon (fig. 3.8). No other likely translation initiation codons were found further downstream. The amino acid sequence of the open reading frame shows no identity with any amino acid sequence in the protein databases searched. It also has no homology with the amino acid sequence of the subregion 2.1 consensus, suggesting that the open reading frame does not encode a sigma factor.

A potential σ^{70}-dependent promoter was identified upstream of the TTG codon. The -10 element, TATACT, has 5 out of 6 homology with the consensus sequence, and point mutations at the non-consensus position 5 of this promoter are known not to be severe down mutations (Youderian *et al.*, 1982). The -35 element, TTTTTT,

has 2 out of 6 homology with the consensus sequence, and is optimally spaced at 17bp upstream of the -10 element. The -35 element overlaps with the proposed rho-independent transcriptional terminator of *fdhF* (Zinoni *et al.,* 1986). In addition, a CCAAT motif present in the promoter region of certain cold-shock genes of *E.coli* was found just downstream of the -10 element. These potential promoter signals are shown in fig. 3.8. DNA sequence analysis of the ORF/*fdhF* intergenic region, and of *fdhF* itself, failed to reveal consensus promoter sequences for alternative sigmas.

Figure 3.6. Dideoxy-nucleotide sequence of the *gltP/fdhF* intergenic region using M13 tg130-4 ssDNA as template (section 2.7.1). Lanes 1-4 show sequence runs primed from within the intergenic region back towards *fdhF*, with the tail of the *fdhF* transcriptional terminator indicated (see fig.3.8). Lanes 5-8 show sequence runs primed downstream from within *gltP* towards the intergenic region using the same DNA template preparation. The dideoxynucleotide used in each sequencing reaction is indicated above the corresponding track.

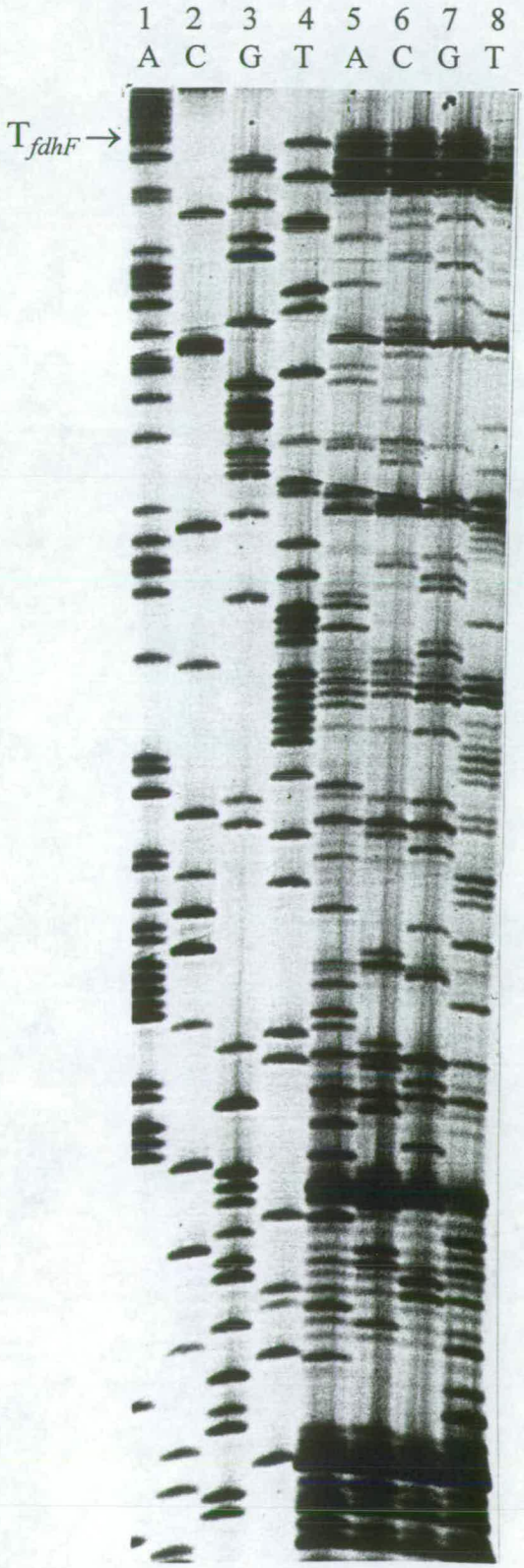


Figure 3.7. Sequencing strategy used for the *gltP/fdhF* intergenic region. Single-stranded M13 tg 130-4 and 131-3 DNA were sequenced as described in section 2.7.1. The top line shows the distance in base pairs, and begins at position 1 of fig. 3.8. The second line gives a simple restriction map of the region. Abbreviations for restriction sites are : V, *Pvu*II; G, *Bgl*II; S, *Stu*I; E, *Eco*RV. Also shown are the *fdhF* terminator (T) and the *f229* promoter (P). Arrows show the direction and approximate extent of sequencing reactions primed with different oligonucleotides, whose numbers are given alongside. Primers shown above the third line were used to sequence tg 131-3; those shown below were used to sequence tg 130-4. Multiple binding sites within the BIME are shown for the primers A866 (A) and C134 (C). The primer M5102 was used for primer extension analysis (section 3.6) and for confirmation of the *Kpn*I/*Eco*NI linker sequence present in pSW11 and other plasmids (fig. 3.14).

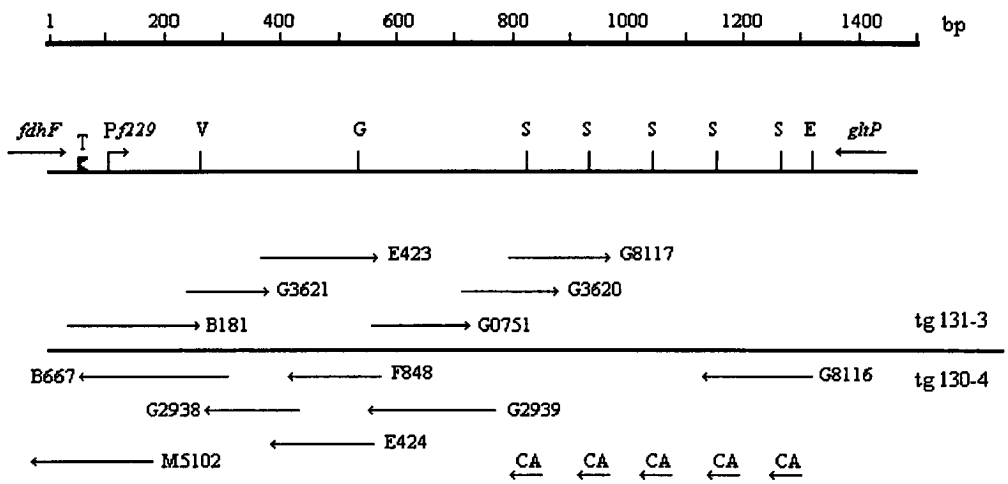
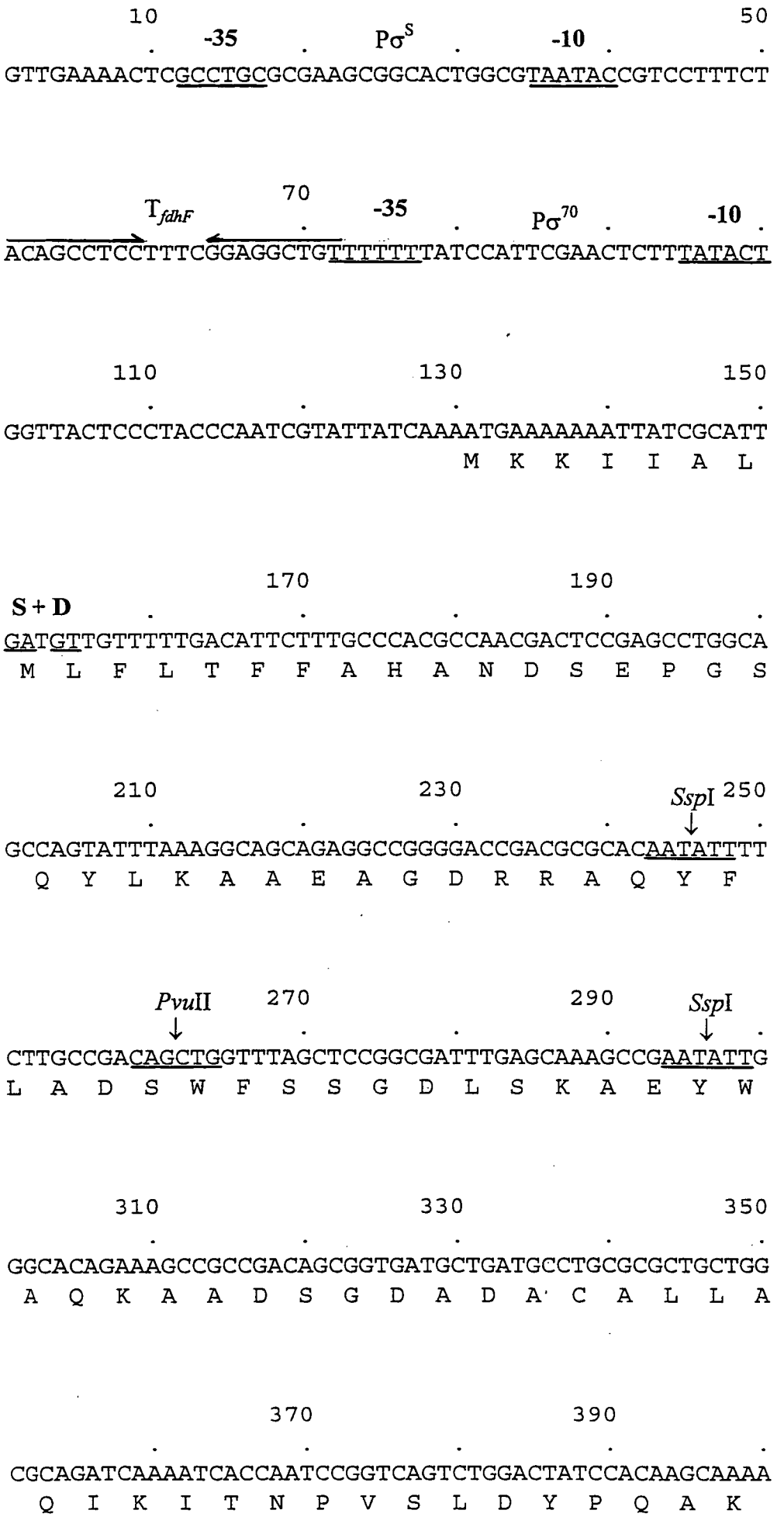


Figure 3.8. Nucleotide sequence of the *fdhF/gltP* intergenic region in the 5′ to 3′ direction of *fdhF* in RNA. The predicted amino acid sequence of the open reading frame is shown below the relevant codons. The putative promoter identified by Blattner *et al.* (1993), the putative transcriptional terminator of *fdhF* identified by Zinoni *et al.* (1986), and the putative promoter identified by the author are indicated. A potential Shine-Dalgarno ribosome binding sequence upstream of a TTG leucine (potential start, F-met) codon within the open reading frame is indicated. Restriction sites mentioned elsewhere in the text are marked. The motifs making up the extragenic repeats are noted. Only the first repeat is shown.


10
50
-35
Pσ^S
-10
GTTGAAAACTCGCCTGCGCGAAGCGGCACTGGCGTAATACCGTCCTTTCT
70
T_fdhF
-35
Pσ^70
-10
ACAGCCTCCTTTCGGAGGCTGTTTTTTTATCCATTCGAACTCTTTATACT
110
130
150
GGTTACTCCCTACCCAATCGTATTATCAAAATGAAAAAAATTATCGCATT
M K K I I A L
170
190
S+D
GATGTTGTTTTTGACATTCTTTGCCCACGCCAACGACTCCGAGCCTGGCA
M L F L T F F A H A N D S E P G S
210
230
SspI
250
GCCAGTATTTAAAGGCAGCAGAGGCCGGGGACCGACGCGCACAATATTTT
Q Y L K A A E A G D R R A Q Y F
PvuII
270
290
SspI
CTTGCCGACAGCTGGTTTAGCTCCGGCGATTTGAGCAAAGCCGAATATTG
L A D S W F S S G D L S K A E Y W
310
330
350
GGCACAGAAAGCCGCCGACAGCGGTGATGCTGATGCCTGCGCGCTGCTGG
A Q K A A D S G D A D A C A L L A
370
390
CGCAGATCAAAATCACCAATCCGGTCAGTCTGGACTATCCACAAGCAAAA
Q I K I T N P V S L D Y P Q A K

GTTCTTGCAGAGAAAGCGGCGCAAGCGGGCAGTAAAGAAGGTGAAGTAAC
V L A E K A A Q A G S K E G E V T

470 490

GCTGGCGCATATTCTGGTAAATACTCAAGCGGGTAAACCGGATTATCCAA
L A H I L V N T Q A G K P D Y P K

510 530 *BglII* 550

AGGCAATTTCGCTGTTAGAAAACGCCTCGGAAGATCTGGAGAACGACTCT
A I S L L E N A S E D L E N D S

570 590

GCCGTCGATGCCCAAATGCTGCTTGGTTTGATTTACGCCAACGGCGTGGG
A V D A Q M L L G L I Y A N G V G

610 630 650

CATTAAGGCCGACGATGACAAGGCAACCTGGTATTTCAAACGCAGCTCTG
I K A D D D K A T W Y F K R S S A

670 *ScaI* 690

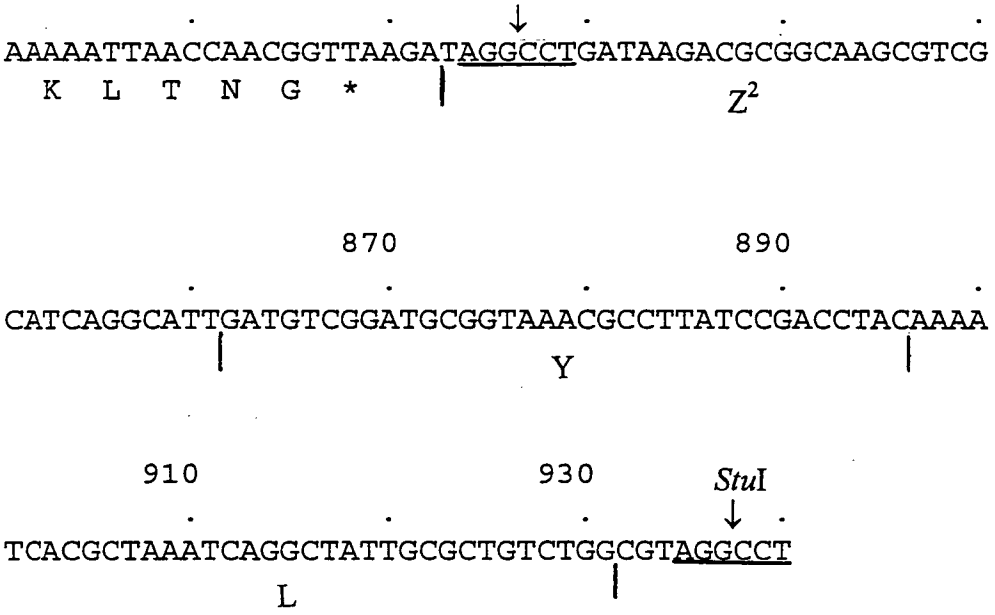
CAATTTCCCGAACCGGTTATTCCGAGTACTGGGCGGGAATGATGTTCTTA
I S R T G Y S E Y W A G M M F L

710 730 750

AACGGTGAAGAAGGTTTTATCGAGAAGAACAAGCAAAAGGCGCTGCACTG
N G E E G F I E K N K Q K A L H W

770 790

GTTGAACCTGAGCTGTATGGAAGGGTTTGATACCGGGTGTGAAGAGTTTG
L N L S C M E G F D T G C E E F E


AAAAATTAACCAACGGTTAAGATAGGCCTGATAAGACGCGGCAAGCGTCG
K L T N G *
Z2
870
890
CATCAGGCATTGATGTCGGATGCGGTAAACGCCTTATCCGACCTACAAAA
Y
910
930
StuI
TCACGCTAAATCAGGCTATTGCGCTGTCTGGCGTAGGCCT
L

3.5 Identification of extragenic repeats

When the sequence downstream of the ORF matched the 3′ end of the published *gltP* sequence, the calculated distance between *fdhF* and *gltP* was found to be smaller than expected. Although RF template DNA gave restriction fragments of the size expected from the Kohara map, some 500bp appeared to remain unsequenced. However, restriction analysis with *Stu*I, which cuts 6bp downstream from the stop codon of the ORF, gave a smaller fragment whose size agreed with that predicted from the sequence. To investigate this further, pSW3 and pSW4 DNA were digested with *Stu*I and analysed by high percentage agarose gel electrophoresis to locate the "missing" DNA. An unexpected band was found to migrate at approximately 110bp, which was more than equimolar with other fragments within that digest, and differed in intensity between pSW3 and pSW4 (fig.3.9a). *Stu*I partial digests finally revealed its whereabouts (fig 3.9b). *Stu*I sites separated by ~110bp and identical sequences at the 3′ ends of *gltP* and the open reading frame suggested a series of repeats in their intergenic region, of which pSW4 has lost 2 relative to pSW3.

While these repeats were being sequenced, the entire genomic sequence between 89.2 and 92.8 minutes was published by Blattner *et al.* (1993) as part of the *E.coli* genome project. The GenBank accession number is U00006; nucleotide number 162495 of this sequence corresponds to nucleotide number 1 of fig. 3.8. Blattner *et al.* also noticed the ORF. They named it *f229* because it is transcribed in the "forward" orientation relative to DNA replication and potentially encodes 229 amino acids. I will refer to the ORF as *f229* from now on. The extragenic region was found to be made up of REP (Repetitive Extragenic Palindromic) sequences, or BIME (Bacterial Interspersed Mosaic Elements) as they are now known (section 1.9). The chromosomal BIME is reported to have six *Stu*I recognition sites. pSW3 as well as pSW4 has a deletion in this highly repetitive region. The BIME is comprised of 5 RIB/RIP elements (section 1.9.5) and one additional PU (section 1.9.1). The sequence and motifs of this PU and the first RIB/RIP proximal to *f229* are shown in

fig. 3.8. The pile-ups observed in fig. 3.6 arose due to multiple primer hybridisation sites within the BIME. A map of pSW3 is given in fig. 3.10.

To further map the probe-positive region pSW3 fragments were probed with 746J. Fig. 3.11a shows that the probe does not hybridise to sequences within the BIME, and along with previous Southern analysis (fig. 3.3) mapped the probe-positive region between the *Pvu*II site of *f229* and the proximal *Stu*I site (fig. 3.8). Dot blots of single stranded tg130-4 and tg131-3 DNA (fig. 3.11b) indicated that 746J hybridises to the coding strand of *f229*, and therefore shares sequence homology with the non-coding strand. This was confirmed by Fasta and Bestfit programs from the GCG package, which found 10 out of 11bp sequence homology between the non-coding strand of *f229* and 746J located approximately 90bp downstream from the PvuII site. 7 of these basepairs are G-C and are found at the middle of the probe :

f229	352-CTGGCGCAGAT-361	(numbered as in fig. 3.8)
746J	54-GTCCGCGTCTA-38	(non-complementary base underlined)

Since mismatches at the centre of a probe have a greater negative effect on hybridisation than those towards the ends (Sambrook *et al.*, 1989) this portion of the probe is thought to form the primary contact with the *f229* target sequence, with additional contacts being made at either side. However, no significant open reading frames have been found in the non-coding strand of *f229*.

Figure 3.9. Restriction analysis of pSW3 and pSW4 DNA.

Fig 3.9A. 3% agarose gel electrophoresis. Track 2 contains 5µg pSW3 DNA digested with *Bgl*II and *Stu*I; Track 3 contains 5µg pSW4 digested with *Bgl*II and *Stu*I. Note that the ~110bp *Stu*I fragment is not equimolar with either the 239bp *Bgl*II fragment or with the 290bp *Bgl*II/*Stu*I fragment. In addition, the same *Stu*I fragment of pSW3 is not equimolar with that of pSW4, although the respective *Bgl*II and *Bgl*II/*Stu*I fragments are. Tracks 1 and 4 contain 5µg 1kb marker (Gibco BRL). The relevant pSW3/pSW4 fragments are marked and their sizes given to the left.

Fig. 3.9B. 0.6% agarose gel electrophoresis. Track 2 contains 5µg pSW3 DNA digested with *Sph*I, then partially with *Stu*I; Track 3 contains 5µg pSW4 DNA digested with *Sph*I, then partially with *Stu*I. Partial digestion was achieved by adding 0.05 Units of *Stu*I in a final volume of 20µl and incubating for 15 minutes at 37°C before terminating the reaction. Track 1 contains 2µg 1kb marker (Gibco BRL). Relevant marker fragment sizes are given in kbp to the left.

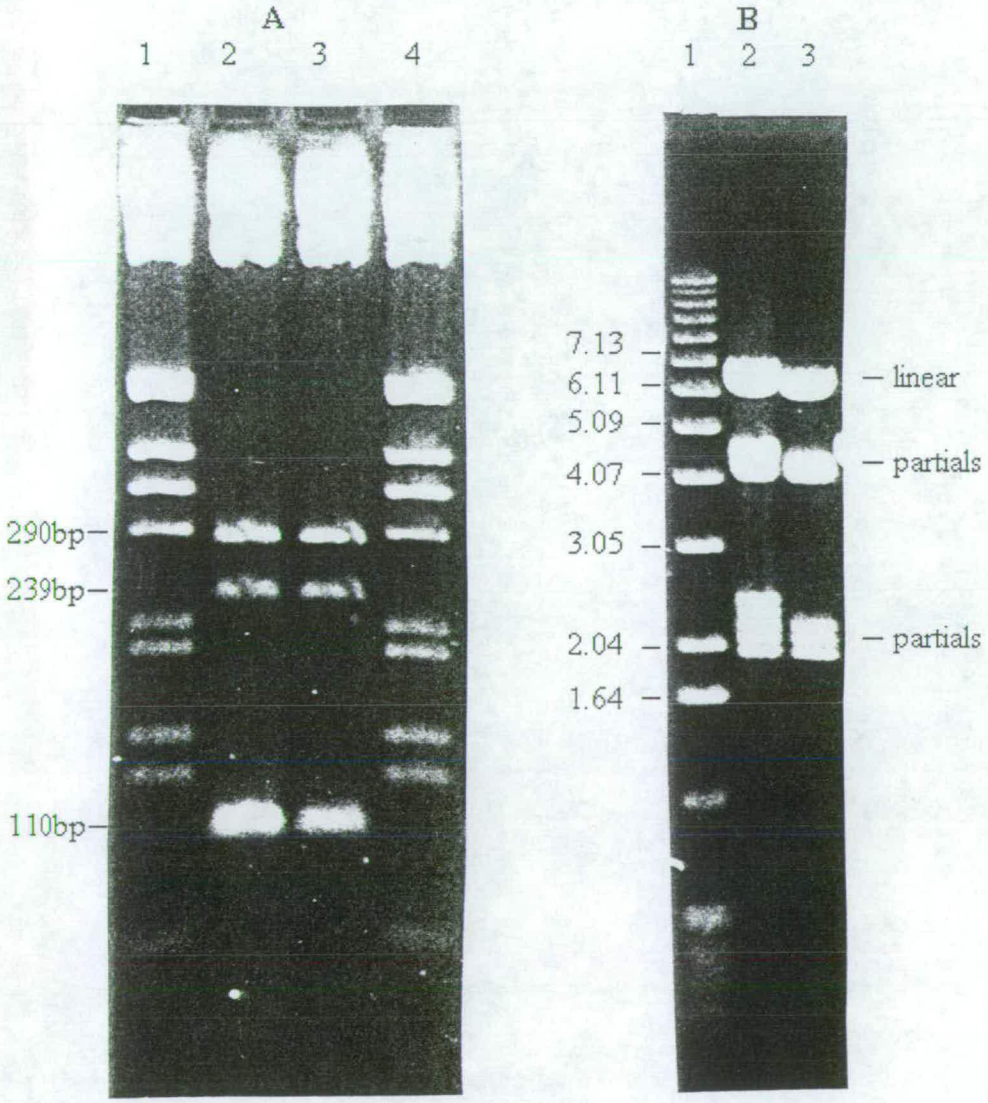
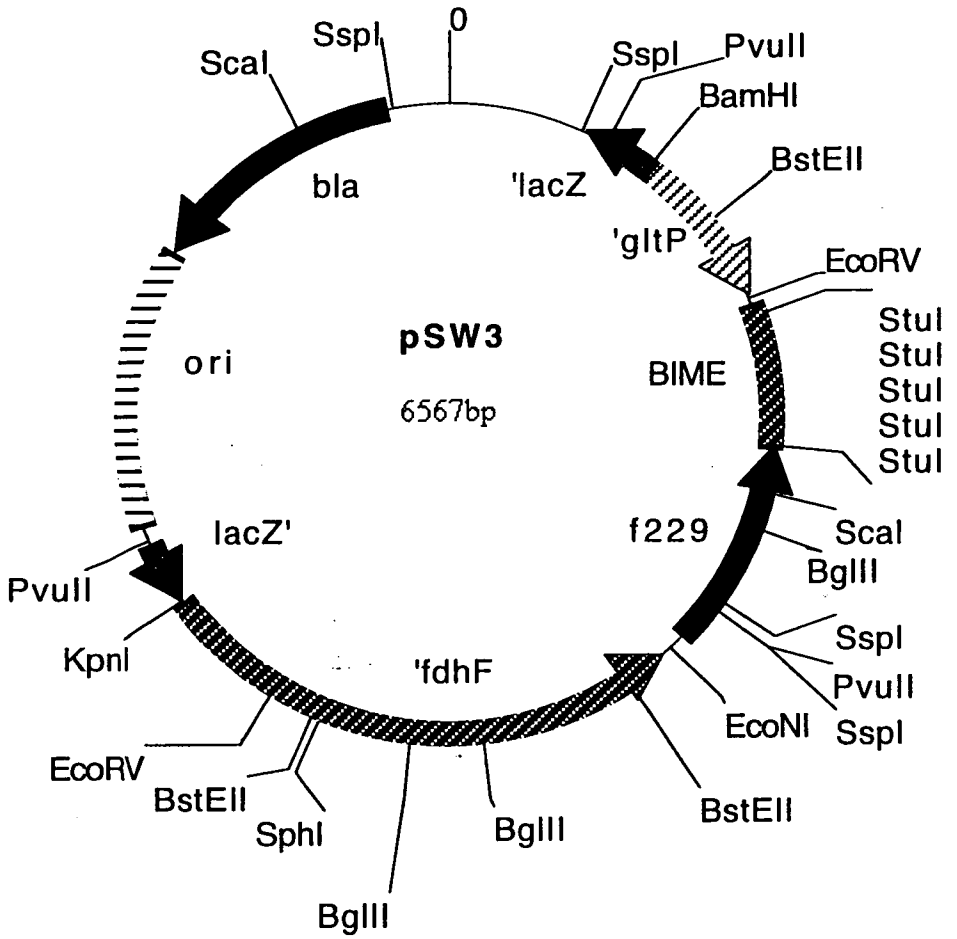


Figure 3.10. Detailed structure of pSW3. A 3.7kb probe-positive *Bam*HI-*Kpn*I fragment from λ639 was subcloned into the corresponding sites in pBluescript KS (-). These restriction sites are marked in fig. 3.4. See text for an explanation of additional features.



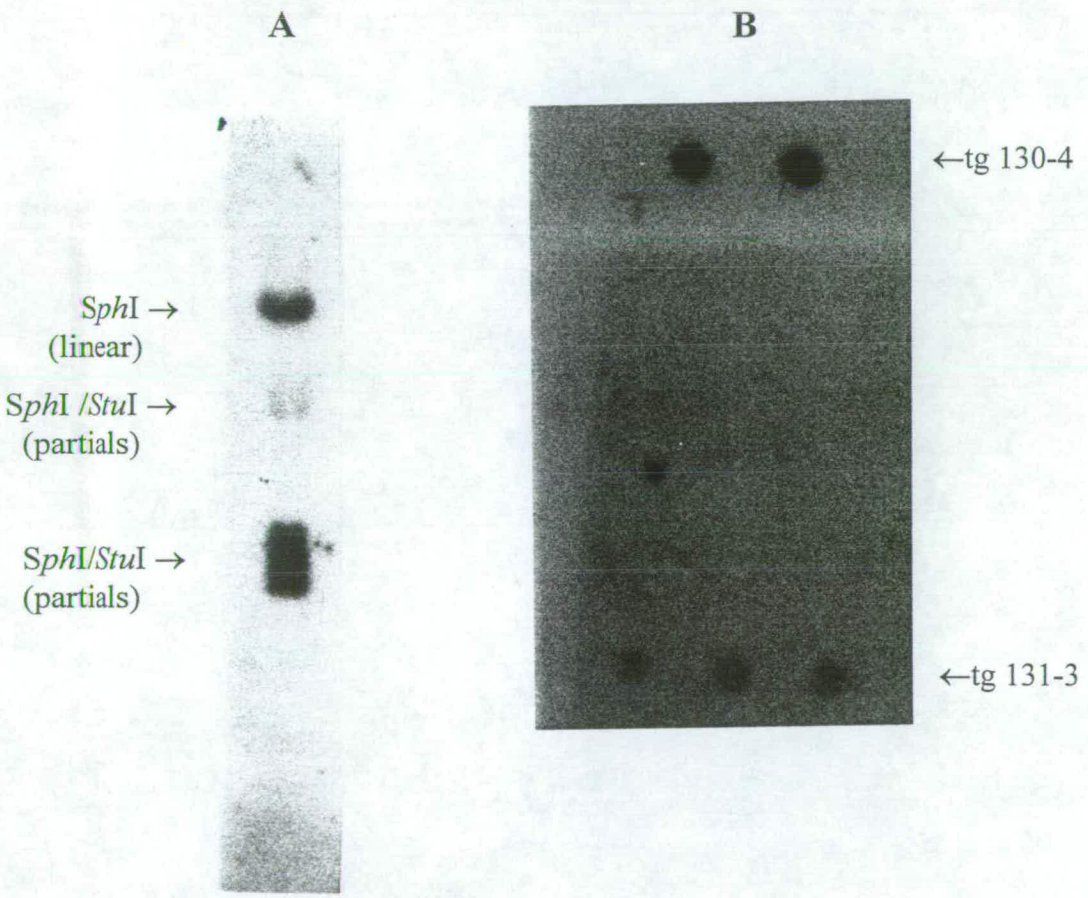
Restriction sites (bp) :

*Bam*HI	689	*Pvu*II	529, 2308, 4513
*Bgl*II	2033, 3177, 3416	*Sca*I	1894, 6067
*Bst*EII	951, 2683, 3717	*Sph*I	3706
*Eco*NI	2506	*Ssp*I	442, 2273, 2325, 6386
*Eco*RV	1239, 3873	*Sst*I	657
*Kpn*I	4295	*Stu*I	1293, 1406, 1519, 1623, 1743

Figure 3.11. Fine mapping of the 746J probe-positive region within *f229*.

Figure 3.11a. Southern hybridisation. pSW3 DNA was digested with *Sph*I then partially with *Stu*I as in fig 3.9b. DNA was transferred to a nitrocellulose membrane as described in section 2.8.3. Hybridisation was performed at 50°C for 18 hours. Final washing conditions were 0.2x SSC, 0.1% SDS for 2 minutes at 65°C.

Figure 3.11b. Dot blot hybridisation analysis of M13 tg130-4 and tg131-3 ssDNA using 746J as probe. 100ng DNA was incubated at 95°C for 5 minutes then spotted onto a nylon membrane as described in section 2.8.3. Hybridisation and washing conditions were as above.



3.6 Transcription analysis of the *f229* gene

The *E.coli* Genome Project sequence analysis data include other features such as possible promoters and terminators and computer-predicted static bends. A potential σ^S-dependent promoter was identified just upstream of the putative rho-independent transcriptional terminator of *fdhF* (fig. 3.8) and a bend of 75° identified in the upstream coding region of *f229* (Blattner *et al.*, 1993). As discussed in section 1.8.1, σ^S-dependent promoters are often associated with bent DNA. However, it should be stressed that even now, σ^S-dependent promoter determinants are not well defined, despite the more recent work of Ishihama, Tanaka and others (section 1.8.1).

3.6.1 *f229* is not detectably transcribed by $E\sigma^S$

To test the possibility that *f229* transcription is under the control of $E\sigma^S$, as suggested by Blattner *et al.* (1993), RNA was extracted from cells grown to various stages of stationary phase and also from cells grown in media containing 80mM sodium acetate, which is known to induce certain σ^S-dependent genes (Mukhopadhyay and Schellhorn, 1994). Both primer extension and Nuclease S1 analysis failed to show induction, even when *f229* was present on the multicopy plasmid pSW3. In addition, *f229* did not appear to be transcribed during exponential growth (data not shown).

Run-off transcription assays with purified $E\sigma^{70}$ and $E\sigma^S$ were also performed using pSW3 promoter fragment templates as described in section 2.9. Again, $E\sigma^S$ failed to give detectable transcripts, whereas $E\sigma^{70}$ produced specific and reproducible bands (fig. 3.12a). From the *lacUV5* marker lane of fig. 3.12a, transcripts of approximately 430nt and 140nt were obtained with promoter templates running off at the *Bgl*II and *Ssp*I restriction sites within *f229* respectively (see also fig.3.12b). Since these restriction sites are separated by 291bp, the transcripts seem clearly to have originated from the same promoter. Their sizes locate the transcription start site in the region of the putative -10 promoter element, TATACT, discussed in section 3.4.

Figure 3.12a. Run-off transcription assay (section 2.9) with $E\sigma^{70}$ (tracks 1-3) and $E\sigma^{S}$ (tracks 4-6) using two different *f229* promoter template fragments from pSW3, and a *lacUV5* promoter template fragment as a control. Tracks 1 and 4 contain transcripts from a 1144bp *Bgl*II fragment, expected to run off at position 532 of the *Bgl*II site in fig. 3.8. Tracks 3 and 6 contain transcripts from a 852bp *Bgl*II/*Ssp*I fragment, which should run off at position 245 of the *Ssp*I site in fig. 3.8.

Tracks 2 and 5 contain transcripts from a 205bp *Eco*RI fragment from pKB252 (Backman *et al.*, 1976) carrying the *lacUV5* promoter. This promoter is recognised by both $E\sigma^{70}$ and $E\sigma^{S}$ (Tanaka *et al.*, 1993). The primary transcript from this promoter fragment is 63nt. In addition, a secondary 268nt transcript and a tertiary 473nt transcript are present. These are thought to be generated by a proportion of RNA polymerase molecules which turn at the end of the template DNA rather than run off, and which then read back along the promoter fragment. The sizes of these marker transcripts are given in nucleotides at the right hand side of the figure. The approximate sizes of the *f229* promoter initiated transcripts are given at the left hand side of the figure. Some of the sample from track 5 has spilled into the preceding track.

Figure 3.12b. Diagram showing the run-off transcripts in tracks 1 and 3 generated by the two *f229* promoter template fragments. The following abbreviations are used for restriction enzymes : G, *Bgl*II; S, *Ssp*I.

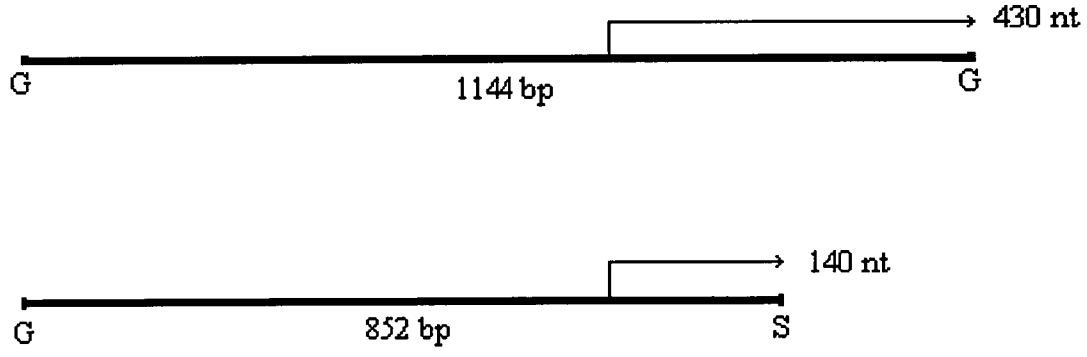
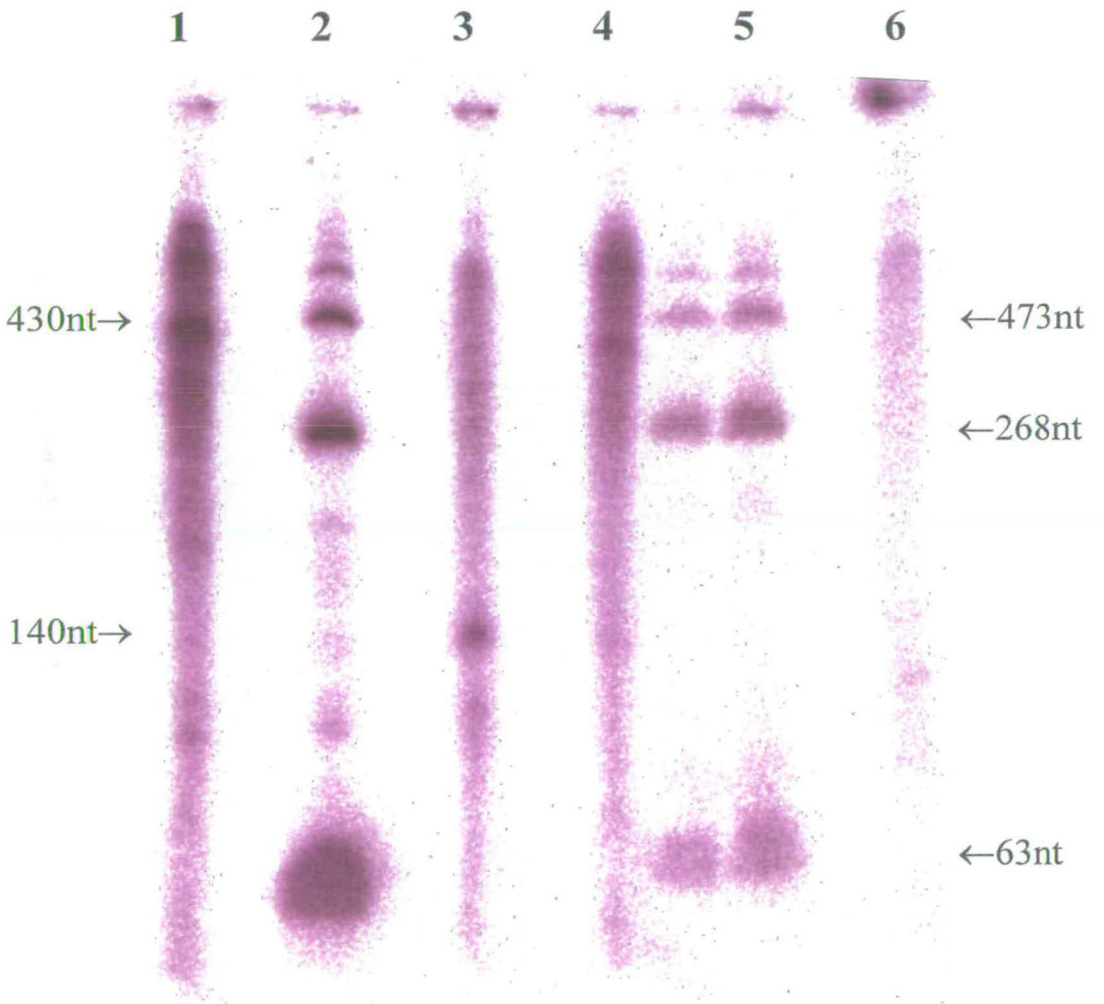


Fig. 3.12a



3.6.2 Mapping the *fdhF* transcriptional terminator

Although the run-off transcription assays indicated that a σ^{70}-dependent promoter is active *in vitro*, no *f229* transcript was detected by primer extension and Nuclease S1 analysis in RNA extracted from exponentially-growing cells (data not shown) suggesting a possible *in vitro* artefact. On the other hand, DNA sequence analysis of the *f229* promoter region failed to reveal other consensus promoter sequences. To test for possible co-transcription of *f229* with *fdhF*, the transcriptional terminator of *fdhF* was mapped using Nuclease S1 as described in section 2.8.9. Although the DNA sequence of *fdhF* clearly suggests the presence of a strong rho-independent terminator, read-through or anti-termination might occur to give low levels of *f229* transcript. Since *fdhF* is transcribed from a σ^{N}-dependent promoter under conditions where formate is the ultimate electron acceptor (Zinoni *et al.*, 1986), transcription of chromosomal *fdhF* is not likely to occur under aerobic growth conditions.

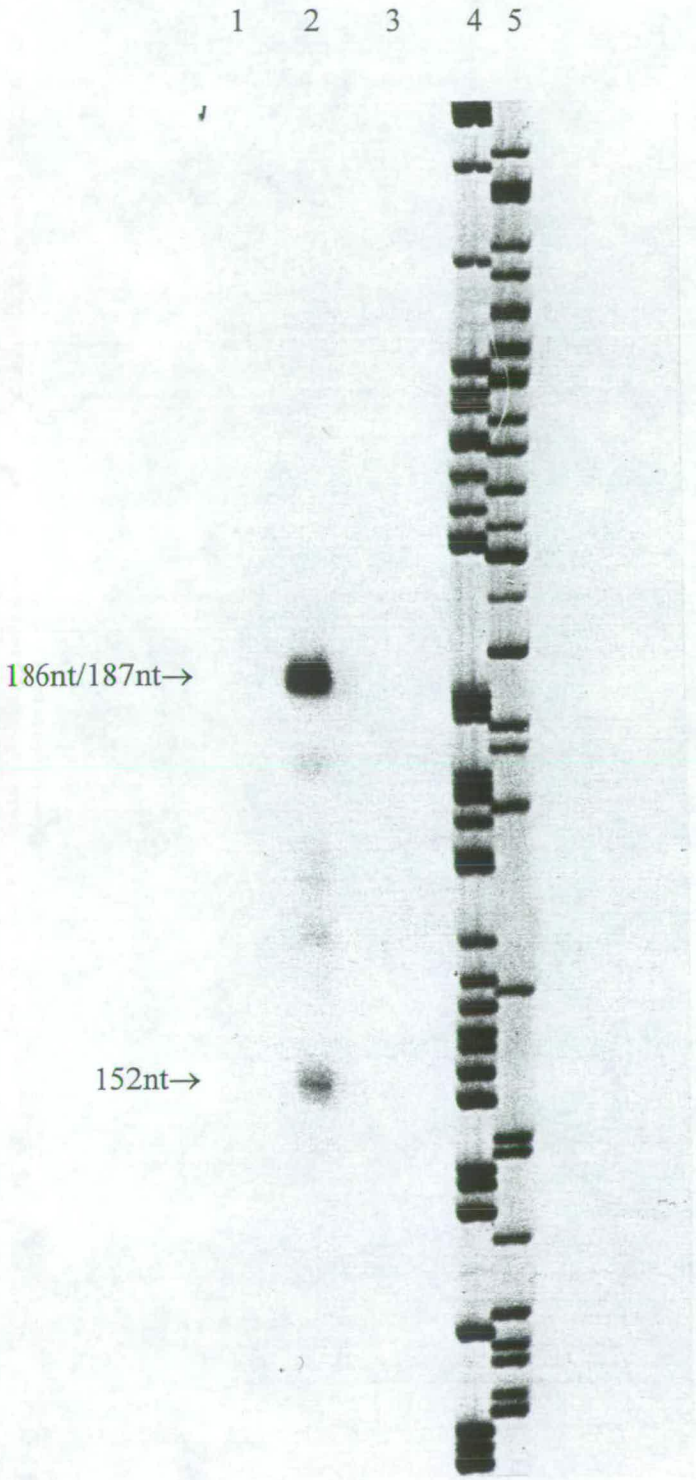
To generate a 3′ portion of *fdhF* mRNA, transcription was initiated from the IPTG-inducible *lac* promoter of pSW3 (fig.3.10). S1 Nuclease analysis consistently gave two bands of 186nt and 187nt (fig. 3.13a). These correspond to the final T of the stem-loop, and the first T distal to this in the 3′-tail of the putative terminator of *fdhF* (starred in fig. 3.13b). Termination would be expected to occur further downstream in the run of Ts. The failure to observe this may well be an artefact arising from the tendancy of Nuclease S1 to "nibble" into A:T (or A:U) basepairs (Sambrook *et al.*, 1989). Alternatively, 3′-exonuclease may trim the mRNA back to the stem-loop *in vivo*. In any case, no full-length uncut probes were detected following S1 digestion. This suggests that there is no detectable read-through of the *fdhF* terminator, and hence no co-transcription of *f229* with *fdhF*, at least when transcription of *fdhF* is initiated from the heterologous *lac* promoter.

Figure 3.13a. Nuclease S1 analysis of the transcriptional terminator of *fdhF*. A 364bp *Bst*EII-*Ssp*I fragment was 3′ end labelled as described in section 2.8.2. This was hybridised with 100μg of total RNA and digested with Nuclease S1 as described in section 2.8.9. Products of Nuclease S1 digestion using the following RNAs were analysed on a 6% denaturing polyacrylamide gel : Track 1, *E.coli* tRNA (Boehringer); Track 2, RNA from NM570.pSW3, induced with 100μg/ml IPTG; Track 3, RNA from NM570.pSW3, not induced with IPTG. Tracks 4 and 5 contain sequencing ladders produced with Sequenase from M13 control DNA using ddA and ddT reactions respectively. The sizes of the major Nuclease S1-generated products are given in nucleotides.
The weak band of 152nt corresponds to position 39 in fig. 3.8, and lies just downstream from the *fdhF* translation stop codon. This band may have been generated by RNase processing or by degradation of the probe.

Figure 3.13b. The nucleotide sequence of the *fdhF* terminator region, showing the inverted repeat (underlined) and T-rich tail, with asterisks above the two Ts corresponding to the 186 and 187nt bands shown in fig. 3.13a. In addition, a direct repeat of TCCTTTC(N)$_7$TCCTTTC and the target sequence for *Eco*NI [CCT(N)$_5$AGG] are also shown. The sequence is numbered as in fig. 3.8.

*Eco*NI
↓ **

43 - TCCTTTCTACAGCCTCCTTTCGGAGGCTGTTTTTTT- 78


1 2 3 4 5
186nt/187nt→
152nt→

3.6.3 *f229* transcription is not detectably cold-shock inducible

13bp downstream of the putative -10 promoter element TATACT lies the complement of the Y-box motif, ATTGG (fig. 3.8). This regulatory sequence is found in the promoter regions of several cold-shock genes of *E.coli* and is bound by the major cold-shock protein CspA (section 1.11.2). No alternative sigma controlling the cold-shock regulon has been identified, and transcription of cold-shock genes may involve activation of $E\sigma^{70}$ (or $E\sigma^{S}$) by cold-shock-induced non-sigma factors. To determine whether *f229* transcription is cold-shock inducible, exponentially-growing cells were shifted from 37°C to 10°C and RNA extracted at various time points following this temperature downshift. Neither primer extension nor S1 Nuclease analysis gave evidence of *f229* transcription at any time point (1-24 hours) following cold-shock, even when cells carried the multicopy plasmid pSW3 (data not shown). Again, the *f229* transcript was also undetectable in RNA samples from exponentially-growing cells at 37°C.

3.6.4 *f229* is strongly transcribed after deletion of the downstream BIME

In an attempt to simplify manipulation of pSW3 by removing "excess" DNA, the upstream *fdhF* DNA was removed. This was achieved by digesting pSW3 with *Kpn*I, the original cloning site from λ639, and *Eco*NI, whose target site lies in the loop of the *fdhF* transcriptional terminator (fig 3.13b). The *fdhF* DNA was then replaced with a *Kpn*I-*Eco*NI linker fragment to give pSW11 (fig. 3.14). The linker was designed to remove the upstream arm of the stem of the *fdhF* terminator, so that transcription of *f229* could be strongly induced from the *lac* promoter if necessary. The natural *fdhF* translational stop codon was retained, since this is in frame with the upstream *lacZα* of pBluescript, and would thus prevent the formation of N-terminal *lacZα* fusions. In addition a direct repeat, $TCCTTTC(N)_7TCCTTTC$, which overlaps the *fdhF* terminator sequence (fig. 3.13b) was also retained since this may represent

an *f229* transcription signal. The linker sequence in pSW11 was confirmed by DNA sequencing as described in section 2.7.

Originally constructed to generate protein from *f229*, the *Bam*HI-*Kpn*I fragment of pSW11 (cloned in pBluescript KS) was subcloned into the multiple cloning site of pBluescript SK, so as to place *f229* downstream of the T7 promoter on the resulting plasmid, pSW12 (fig. 3.14). The protein work is described in section 3.7. To investigate the effect of the BIME on *f229* transcription, pSW12 was digested with *Stu*I so as to remove all but one PU of the BIME structure to give pSW15 (fig. 3.14). While primer extension analysis of exponential phase RNA extracted from NM570 cells carrying pSW12 gave the expected negative result, pSW15 RNA gave a very strong signal (fig. 3.15). Primer extension analysis was then repeated with pSW15 RNA and run alongside a pSW3 DNA sequencing ladder to map the transcription initiation start point. Fig. 3.16a shows transcription originating at a T located 7bp downstream from the TATACT sequence, confirming this as the -10 element of the σ^{70}-dependent *f229* promoter. Fig. 3.16b shows the promoter nucleotide sequence with the transcription initiation site indicated. Transcripts originating here will carry the CCAAT Y-box motif near their 5′ ends. However, additional induction of *f229* from pSW15 following cold-shock was not investigated, since the primary aim was to identify conditions which induce *f229* transcription with the full BIME sequence present downstream.

Primer extension analysis was also performed on RNA extracted from exponential phase cells carrying pSW13 (grown in the presence of 1% glucose to keep its wild type *lac* promoter catabolite repressed). pSW13 carries an *f229*::*lacZ*α fusion which has none of the BIME motifs (section 3.7.3 and fig. 3.14). This analysis consistently detected a major transcript initiating at the same position as seen with pSW15 in fig. 3.16 (data not shown).

Figure 3.14a. Structures of pSW9, pSW10, pSW11, pSW12, pSW13, and pSW15.

The *fdhF* DNA of pSW3 (fig. 3.10) was replaced with a *Kpn*I/*Eco*NI linker carrying a translational stop codon in frame with the upstream *lacZ*α (see below) to generate pSW11. pSW9 is identical to pSW11 except for an internal *Stu*I deletion, leaving just one PU of the BIME sequence. pSW10 was formed by inserting a Kanamycin cassette from pUC4-KISS into the unique *Bgl*II site of pSW9, which lies within *f229*. *Bam*HI cuts within the KanR gene.

pSW12 was formed by subcloning the 1.8kb *Kpn*I/*Bam*HI fragment of pSW11 into the corresponding sites of pBluescript SK. This plasmid carries *f229* and the BIME sequence downstream from the T7 promoter. pSW15 is identical to pSW12 except for an internal *Stu*I deletion, leaving just one PU of the BIME.

pSW13 was formed by digesting pSW11 with *Bgl*II and *Bam*HI and religated to give an in-frame fusion between *f229* and *lacZ*α.

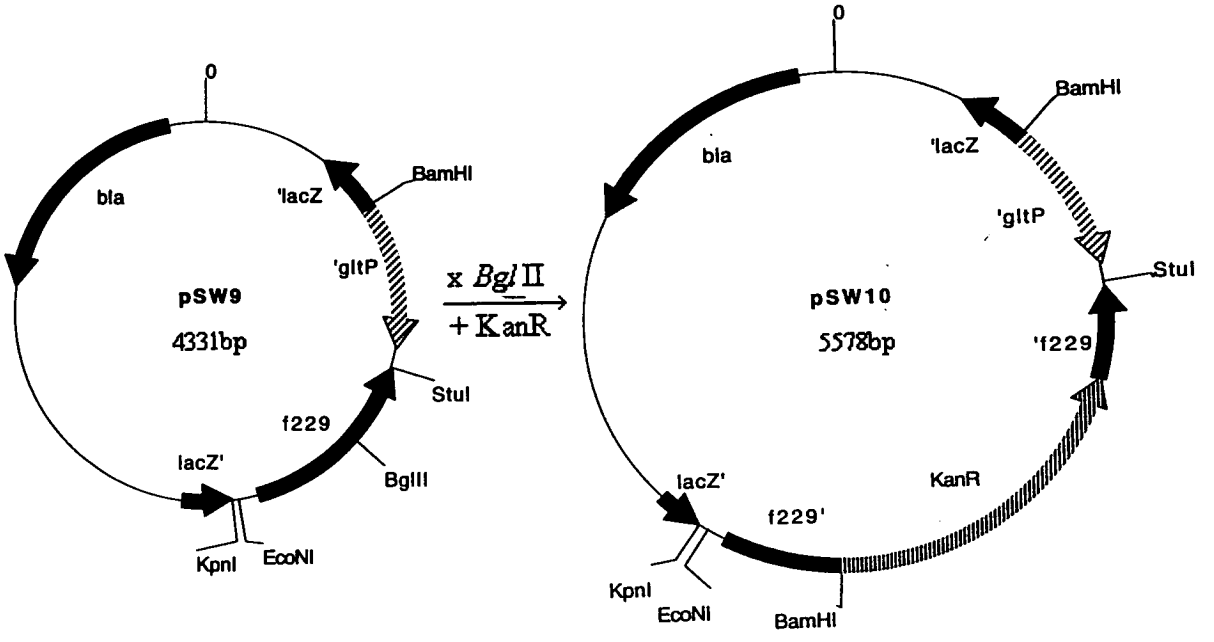


Figure 3.14b. Nucleotide sequence of the *Kpn*I/*Eco*NI linker oligonucleotides. The top line gives the sequence (5′-3′) of oligonucleotide P5335; the bottom line gives the sequence (3′-5′) of oligonucleotide P5336. The TAA translational stop codon is shown in bold.

*Kpn*I *Eco*NI

CGT**AA**TACCGTCC TTT CGCACTT GTCC TTT

CATGGCA**TT**ATGGCAGGAAAGCGTGAACAGGAAAG

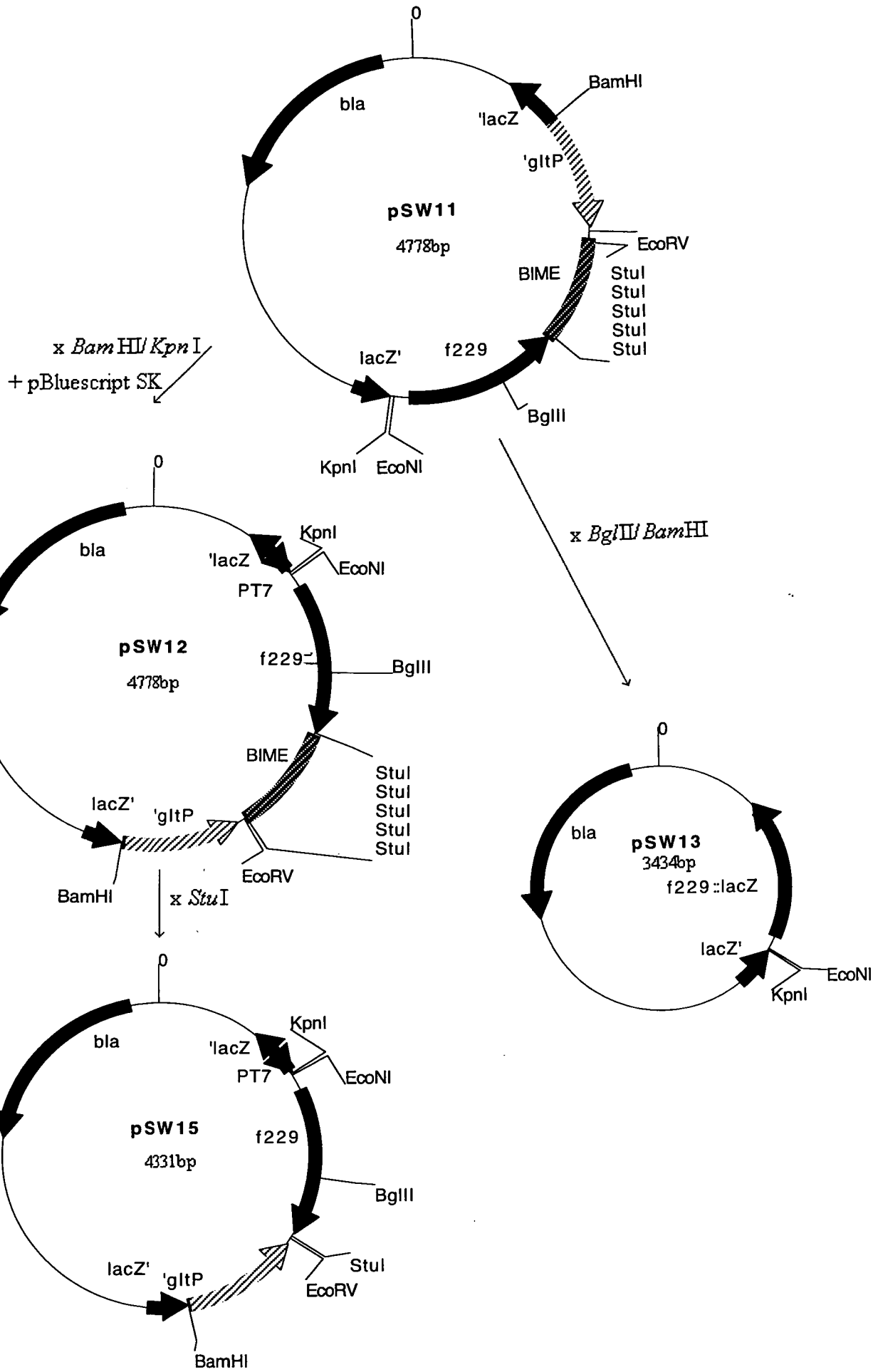

0
bla
'lacZ
BamHI
'gltP
pSW11
4778bp
EcoRV
BIME
StuI
StuI
StuI
StuI
StuI
lacZ'
f229
BglII
KpnI
EcoNI
x BamHI/KpnI
+ pBluescript SK
x BglII/BamHI
0
bla
'lacZ
KpnI
EcoNI
PT7
pSW12
4778bp
f229
BglII
BIME
StuI
StuI
StuI
StuI
StuI
lacZ'
'gltP
BamHI
EcoRV
x StuI
0
bla
pSW13
3434bp
f229::lacZ
lacZ'
EcoNI
KpnI
0
bla
'lacZ
KpnI
EcoNI
PT7
pSW15
4331bp
f229
BglII
lacZ'
'gltP
StuI
EcoRV
BamHI

Figure 3.15. Primer extension analysis of *f229* transcripts. 50μg of RNA from NM570 transformed with pSW12 or pSW15 were hybridised with 10 pmol [γ-^{32}P]-labelled oligonucleotide primer M5102 (complementary to positions 207-181 of fig. 3.8) at 50°C, and primer extended with reverse transcriptase as described in section 2.8.8.

Track 1 shows the sequencing ladder from the Sequenase M13 control ddC reaction; Track 2 shows reverse transcripts generated from pSW15 mRNA; Track 3 shows reverse transcripts generated from pSW12 mRNA. The weaker bands less than 101nt might have arisen due to premature termination of reverse transcription within the mRNA template. It is possible that these are within regions of strong secondary structure, or represent processing sites.

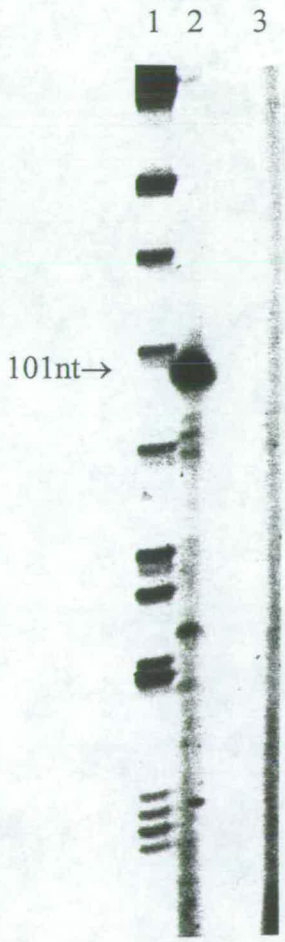
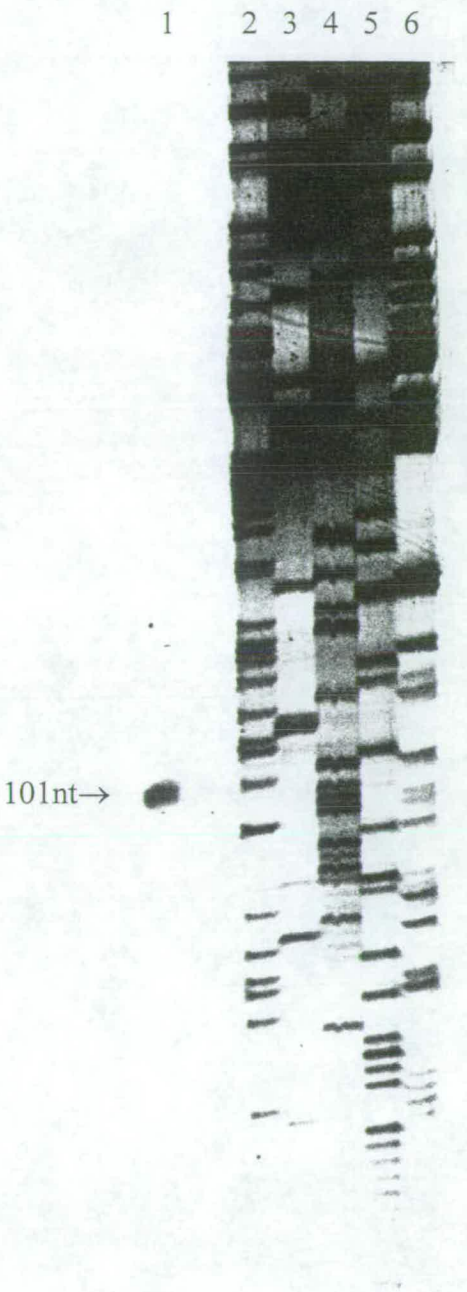


Figure 3.16a. Identification of the *f229* promoter. 50μg of RNA from NM570.pSW15 were hybridised with 10pmol M5102 and extended as in fig. 3.15. Track 1 contains reverse transcripts generated from the above primer extension. Tracks 2-5 contain dideoxy sequencing reactions using CsCl-purified double stranded pSW3 DNA with ddA, ddC, ddG, ddT respectively. The 101nt reverse transcript (at position 107 in fig. 3.8) is marked. Track 6 contains the Sequenase M13 control DNA ddC reaction.

Figure 3.16b. Nucleotide sequence of the *f229* promoter, showing the -35 and -10 elements and the Y-box motif. The transcription start site is indicated with an asterisk.

-35 $(N)_{17}$ -10 * Y-B

TTTTTT TATCCATTCGAACTCTT TATACT GGTTAC TCCCTAC CCAAT


1 2 3 4 5 6
101nt→

3.6.5 *f229* transcription is not induced by osmotic shock

The effects of removal of the BIME indicate that this element exerts a negative effect on *f229* transcription, probably by supercoiling the *f229* promoter DNA. Transcription of osmotic-shock genes is thought to be induced by increased levels of supercoiling, whereas entry into stationary phase results in a general decrease (section 1.10.2). Since stationary phase growth is unable to induce *f229* transcription (section 3.6.1), the level of supercoiling of the *f229* promoter putatively required to allow initiation would have to be even lower than that found in stationary phase. Alternatively, *f229* transcription might for example require the production of an activator which binds to highly supercoiled DNA templates. To test for induction of *f229* transcription by osmotic shock, RNA was extracted from cells grown in LB media containing 0.5M NaCl final concentration (as described in section 2.5.7) to increase the osmolarity of the growth medium. However, RNA dot blots probed with an oligonucleotide complementary to the *f229* transcript show no more hybridisation than RNA from cells grown in normal LB medium (data not shown).

3.7 Analysis of *f229* translation

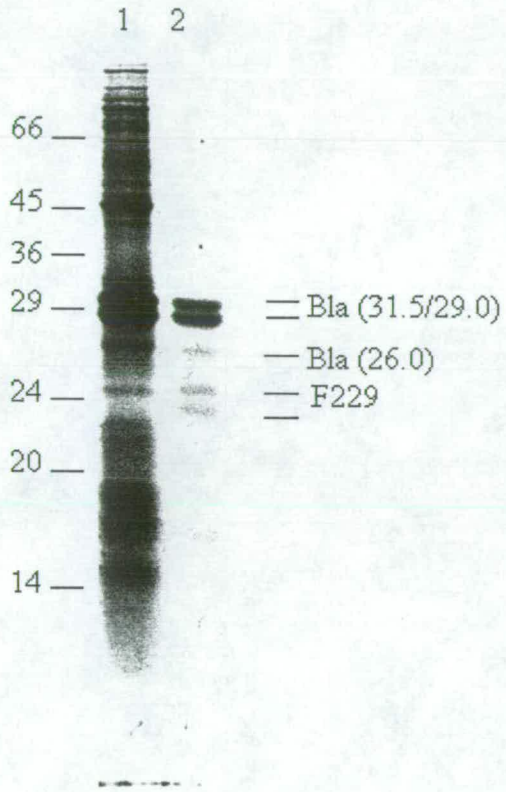
Although *f229* is an open reading frame, no evidence had yet been presented that it is translated, and so its product F229 has remained a hypothetical protein. Results presented in this section provide evidence that *f229* does encode a protein.

3.7.1 Minicell experiments

From its amino acid sequence, *f229* is predicted to encode a protein of 25.1 kD, assuming initiation at the first methionine of the open reading frame as suggested by Blattner *et al.,* 1993). However, proteins labelled in minicells (section 2.10.1) carrying pSW3 and pSW4 plasmids showed no detectable band of this size on SDS-PAGE (data not shown). In the light of the transcription experiments, this could be explained by transcriptional silencing of *f229* due to the BIME present on these plasmids. Accordingly, minicells carrying pSW15, which does give detectable transcripts, consistently gave a protein band migrating at approximately 25kDa as seen by autoradiography (fig. 3.17). A smaller protein band migrating at approximately 23kDa was also detected in the pSW15-generated sample (fig.3.17). This might be a degradation product of F229, or perhaps a truncated F229 protein generated following the cleavage of an N-terminal signal sequence (see section 3.9.6 for further discussion).

Despite producing high levels of transcript (section 3.6.4), pSW13 did not produce the expected ~20kDa *f229::lacZ*α fusion protein in minicells (data not shown).

Figure 3.17. SDS-PAGE analysis of labelled proteins produced by the minicell strain P678.54 carrying either pBluescript SK (track 1) or pSW15 (track 2). Minicells were purified and proteins labelled as described in section 2.10.1. Protein molecular sizes are given in kDa to the left of track 1. The *bla* gene products and F229 are marked in track 2.



3.7.2 A T7 expression system

In view of the minicell results using pSW3 and pSW4 (which were obtained much earlier that those using pSW15) it was logical to attempt over-expression of F229. To achieve this, *Bam*HI/*Kpn*I fragments from pSW3 and pSW4 were first subcloned into M13 tg131 RF DNA to pick up extra restriction sites. *Hind*II/*Sma*I fragments from these M13 intermediates were then transferred into the T7 expression vector pET3 to give pSW5 and pSW6 (see fig. 3.18). The structures of pSW5 and pSW6 were verified by restriction analysis. Cells harbouring these plasmids were infected with M13 mGP1-2, which carries the T7 RNA polymerase gene, and protein synthesis was induced as described in section 2.10.2. Extracts analysed by SDS-PAGE clearly show a protein encoded by pSW5 and pSW6 that is not encoded by the parent plasmid pET3 (fig. 3.19). The size of this protein, which migrates just below the 26kDa degradation product of β-lactamase, is consistent with the predicted size of 25.1kDa for F229. *E.coli* strain BL21(λDE3), which carries the T7 RNA polymerase gene on a λ prophage (Studier and Moffatt, 1986), gave a similar sized band when transformed with pSW12 (data not shown).

It was hoped that this system would generate sufficient levels of protein so that it could be purified. It may have then been possible to confirm by N-terminal sequencing that the ~25kDa protein was indeed F229. However, as with the minicell experiments, the ~25kDa protein band was not seen with Coomassie staining alone (data not shown). Silver staining was not attempted.

Figure 3.18. Structure of pSW5. The T7 promoter lies upstream of *'fdhF, f229,* and *bla.* pSW6 is derived from pSW4 and thus has two fewer repeat units within the BIME than pSW5.

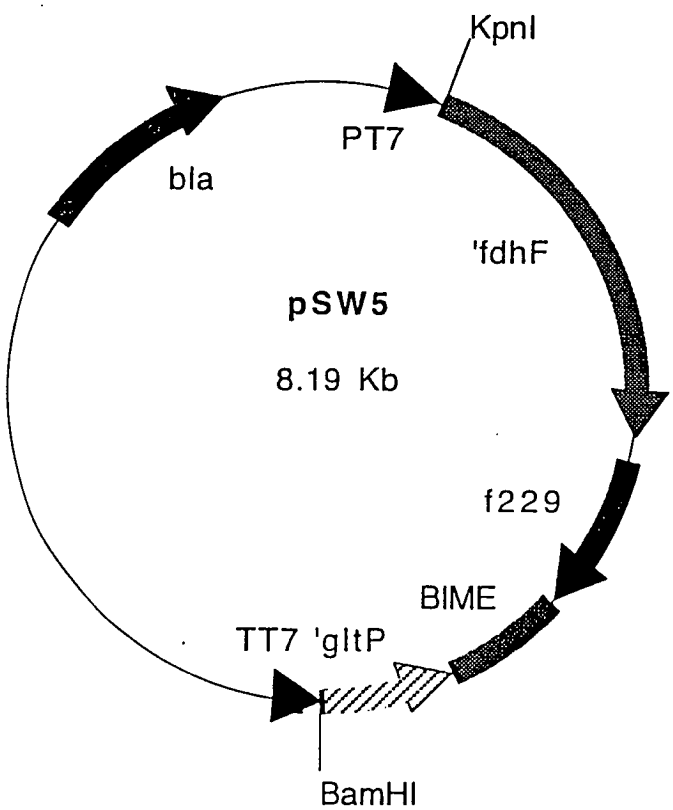
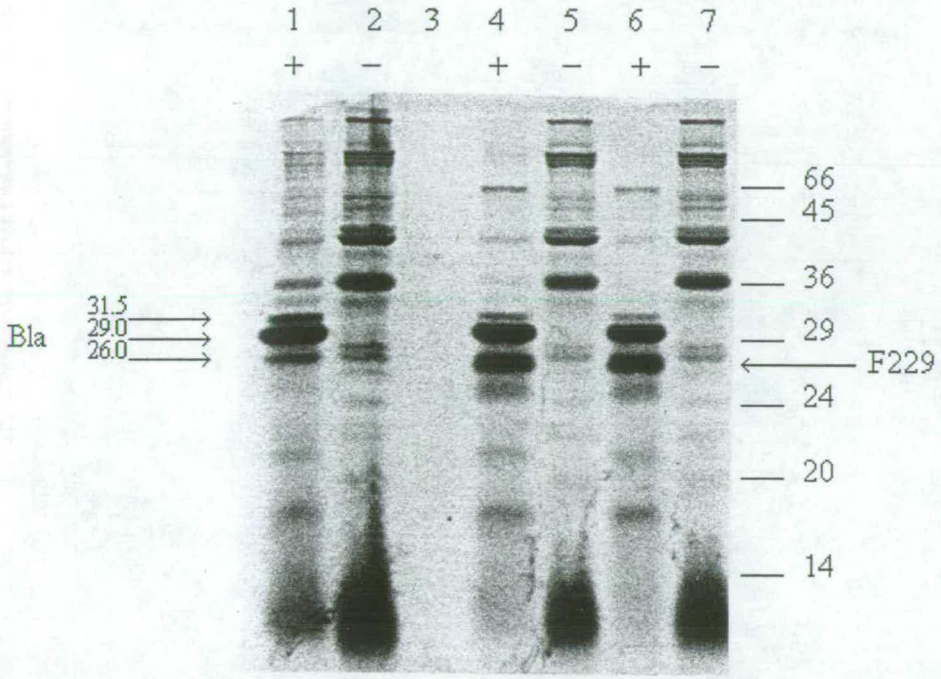


Figure 3.19. SDS-PAGE analysis of labelled proteins produced by pSW5 and pSW6 following induction of transcription by T7 RNA polymerase (section 2.10.2).

+ indicates the addition of both mGP1-2 and IPTG. - indicates addition of neither. Tracks 1 and 2 contain proteins from pET3-carrying cells; Track 3 contains protein molecular size marker (Sigma); Tracks 4 and 5 contain proteins from pSW5-carrying cells; Tracks 6 and 7 contain proteins from pSW6-carrying cells. Protein molecular sizes are given in kDa. The *bla* gene products are marked next to track 1 and F229 is arrowed in lane 6.



3.7.3 *f229* has a functional ribosome binding site

Further evidence that *f229* is translated was obtained using an *f229*::*lacZα* fusion which was constructed by digesting pSW11 with *Bgl*II and *Bam*HI and religating the complementary overhangs to give pSW13 (fig. 3.14). This should produce an in-frame fusion of *f229* and *lacZα*, whereas filling in both ends with Klenow DNA polymerase I fragment (as section 2.8.1 but with unlabelled 2mM dCTP) before religating, which generates pSW14, results in an out of frame fusion. NM570 transformed with pSW13 gave blue colonies on LB plates containing IPTG and X-Gal (section 2.4.6), whereas NM570.pSW14 colonies were white (data not shown). This shows that the protein fusion of pSW13 is able to give α-complementation, and since a translational stop codon in the linker is in frame with the upstream portion of *lacZα*, translation of this fusion protein requires the *f229* ribosome-binding site. pSW13 produced no detectable increase in β-galactosidase activity when assayed by the standard Miller technique (data not shown) although the X-Gal test is more sensitive.

3.7.4 Codon usage in *f229*

The results obtained from the T7 expression system suggest that *f229* might be poorly translated. A possible explanation for inefficient translation is a requirement for rare tRNA species which might be sequestered by more abundant transcripts. The codon usage of the *f229* gene was analysed using the Codonfrequency program of the GCG package, and is summarised in table 3.2. It does indeed show a poor correlation between codon usage and the major iso-accepting tRNA species of *E.coli* (Ikemura, 1981 and 1985). The frequency of optimal codons (F_{op}) was calculated by dividing the number of optimal codons by the total number of optimal plus non-optimal codons assigned according to Ikemura (1985). The calculated low F_{op} (0.52) is consistent with the observation that *f229* is poorly expressed.

Table 3.2 Codon usage in *f229*

First Letter	Second Letter: U			C			A			G		
	UUU	Phe	7	UCU	Ser	2	UAU	Tyr	7	UGU	Cys	2
	UUC	Phe	3	UCC	Ser	4	**UAC**	Tyr	2	UGC	Cys	1
U	UUA	Leu	4	UCA	Ser	0	UAA	Ochre	1	UGA	Opal	0
	UUG	Leu	6	UCG	Ser	2	UAG	Amber	0	UGG	Trp	5
	CUU	Leu	3	CCU	Pro	1	CAU	His	1	**CGU**	Arg	0
	CUC	Leu	0	CCC	Pro	0	CAC	His	2	**CGC**	Arg	2
C	CUA	Leu	0	CCA	Pro	2	CAA	Gln	6	CGA	Arg	2
	CUG	Leu	10	**CCG**	Pro	2	**CAG**	Gln	3	CGG	Arg	0
	AUU	Ile	6	**ACU**	Thr	1	AAU	Asn	2	AGU	Ser	2
	AUC	Ile	4	**ACC**	Thr	5	AAC	Asn	8	AGC	Ser	7
A	AUA	Ile	0	ACA	Thr	1	AAA	Lys	11	AGA	Arg	0
	AUG	Met	6	ACG	Thr	1	AAG	Lys	7	AGG	Arg	0
	GUU	Val	1	**GCU**	Ala	1	GAU	Asp	8	**GGU**	Gly	7
	GUC	Val	2	GCC	Ala	13	GAC	Asp	8	GGC	Gly	5
G	**GUA**	Val	2	**GCA**	Ala	10	**GAA**	Glu	10	GGA	Gly	2
	GUG	Val	1	**GCG**	Ala	9	GAG	Glu	7	GGG	Gly	3

$$\mathbf{F_{op} = 0.52}$$

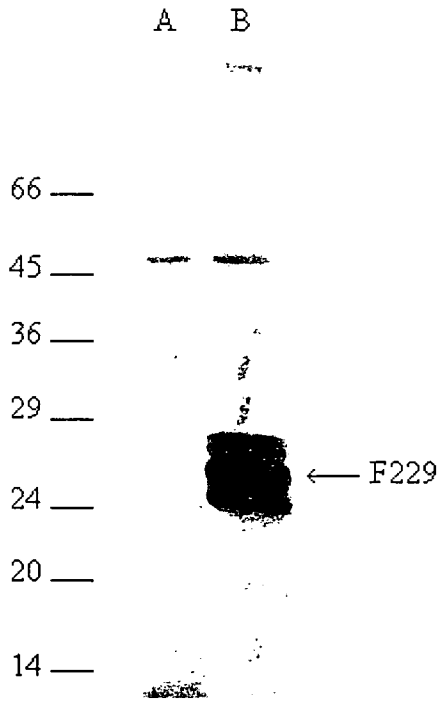
Bold type denotes optimal codons for *E.coli* as designated by Ikemura (1985).

3.7.5 *In vitro* translation of *f229*

F229 production *in vitro* was investigated using a linked T7 transcription/translation system supplied by Amersham, as described in section 2.10.3. Fig 3.20 shows proteins generated from a pSW15 DNA template linearised at either the *Eco*RV or *Bgl*II restriction site (see fig. 3.14). The strong ~25kDa F229 band derived from the *Eco*RV template is indicated; the smaller band might be identical to that seen in minicell samples (fig. 3.17). The identities of the two larger proteins are unknown, although the same banding pattern was observed using pSW15 DNA templates linearised at either the *Stu*I or *Bam*HI site, or with ccc pSW15 DNA (data not shown).

The pSW15 template linearised at the *Bgl*II site would be expected to give a truncated protein of approximately 14kDa. However, no protein of this size was detected, either on gels run for a shorter period, or on overexposed autorads (data not shown). Further inspection of the *f229* sequence showed no methionine codons lying between the *Bgl*II site and the putative AUG initiation codon (Blattner *et al.*, 1993; fig. 3.8). Likewise, the ~20kDa *f229*::*lacZ*α fusion protein encoded by pSW13 has no internal methionine codons. This fusion protein was not seen in minicell experiments (section 3.7.1). Perhaps the N-terminal f-Methionine encoded by the potential UUG initiation codon (or the upstream AUG, if active) is cleaved off the protein. Alternatively, both of these proteins might be highly unstable.

Figure 3.20. SDS-PAGE analysis of proteins labelled in a linked *in vitro* transcription/ translation system (section 2.10.3) with linear pSW15 DNA templates. Track 1 contains proteins from pSW15 linearised at the *Bgl*II site within *f229*; Track 2 contains proteins from pSW15 linearised at the *Eco*RV site downstream from the BIME (see fig. 3.14). The sizes of protein markers (Sigma) are given.



3.7.6 *f229* is not preferentially translated following cold-shock

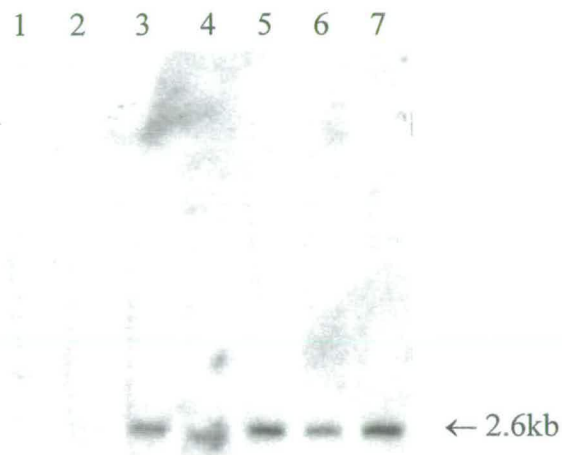
Considering the low F_{op} value for *f229*, its protein product might only increase significantly, if at all, under conditions where translation is specifically directed towards F229 expression. During the lag phase following cold-shock, most cellular protein synthesis ceases except that of cold-shock proteins. In addition, ribosomes appear to be modified so as to preferentially translate the *cspA* transcript, although it is not yet known if translation of other cold-shock transcripts is also enhanced (section 1.11.1). Transcription of *f229* is not induced by cold-shock (section 3.6.3), but translation of pre-existing *f229* transcript (which carries the CCAAT Y-box motif) might be stimulated following cold-shock due to the increased availability of minor tRNAs and/or ribosome modification. However, analysis of proteins extracted from pSW15-carrying minicells labelled at 10°C for between 1 and 4 hours failed to indicate cold-shock-induced translation of *f229* (data not shown).

3.8 *f229* is inessential for growth under conditions studied

The chromosomal copy of *f229* was disrupted in strain NM621 with a kanamycin cassette using pSW10 (fig.3.14) as described in section 2.4.5. Candidates were selected for resistance to kanamycin and sensitivity to ampicillin, and then DNA minipreps were made to ensure no plasmid contamination. *f229* disruption was verified by probing genomic DNA with a kanamycin probe (fig. 3.21). Growth of the *f229*::Kan^R strain was unaffected under various conditions, including cold-shock, osmotic-shock, and stationary phase growth (data not shown). This is perhaps unsurprising considering the observed lack of elevated transcription in response to these stresses.

Figure 3.21. Verification of *f229* disruption on the *E.coli* K12 chromosome. Genomic DNA was prepared as described in section 2.5.6 and was transferred to a nylon membrane by Southern blotting (section 2.8.3). A double stranded DNA fragment carrying the gene encoding aminoglycoside 3′-phosphotransferase, which confers resistance to kanamycin, was labelled as in section 2.8.1 and used as a probe.

Track 1 contains 1 kb ladder marker DNA (Gibco BRL); Track 2 contains NM621 genomic DNA digested with *Bam*HI; Tracks 3-7 contain transductant DNAs digested with *Bam*HI. The inserted gene is present on the chromosome within a 2.6kb *Bam*HI fragment (*Bam*HI cuts in this gene-see fig. 3.14).



3.9 Discussion

3.9.1 The *f229/gltP* BIME

The BIME separating *f229* and *gltP* on the *E.coli* chromosome is comprised of 5 RIB/RIP elements and one additional PU (*gltP* $[Z^2$ L Y s$]_5$ Z^2 *f229*) (section 1.9). The eleven PU sequences which make up this BIME are the greatest number known in one BIME as determined from the sequence data available. There is one known occurrence of a Z^2-containing BIME having a total of ten PU motifs (Bachellier *et al.*, 1994), located between the *phnA* and *phnB* genes of the *phn* operon involved in alkylphosphonate uptake. However, these PUs do not make up RIB/RIP elements. An oligonucleotide having the conserved PU sequences gave the strongest signal with λ639/λ640 when used to probe the Kohara library of *E.coli* genome fragments (Stern *et al.,* 1984). This could mean that the *gltP/f229* BIME carries the greatest number of PU sequences on the chromosome. However, the use of a consensus probe rather than a mixture of three specific probes directed against the Y, Z^1 and Z^2 PU motifs may only reveal a subpopulation of PU, leading to an underestimation of their number (Dimri *et al.,* 1992).

Although the Y and L motifs are repeated with almost total fidelity throughout the BIME, the loop sequence of the Z^2 motif alternates thus : TTGCC, TTAAC, TTGCC, TTAAC, TTAAC, TTGCC. This is uncommon, since the loop sequence is thought to be highly conserved within a particular class (Bachellier *et al.* 1994). In addition, the PU motif is more conserved within a BIME than between BIMEs, leading to the theory that BIMEs are generated by a localized amplification of a PU doublet (Gilson *et al.* 1991). This may still be the case here, since the sequences only differ at two positions, or alternatively, the mechanism of BIME formation may be more complex than at first thought.

The Y motif proximal to the end of *f229* has a TG deletion compared to the other four Y motifs, but this would not affect base pairing. The s motif following this has a single base pair substitution, but as no direct role has been suggested for s, this

may not affect overall function. Comparison of BIME motif sequences indicates that differences at such non-critical positions are quite common.

81% of L motifs carrying the IHF-binding sequence are associated with the Z^1 motif, while 89% of Z^1 motifs are in BIMEs containing just two PU motifs (Bachellier *et al.* 1994). In most cases, BIMEs carrying the L motif have a single copy of L , Z^1 and Y, which are RIBs/RIPs and belong to the BIME-1 family. However, the *glt*P/ *f229* BIME carries five copies of L, each of which is associated with a Z^2 motif. It appears that this BIME structure is rare if not unique, suggesting that it might have a function distinct from other BIMEs. Possibilities include a direct involvement in the attachment of independent chromosomal domains to each other or to the membrane, or in chromosomal segregation.

Since gyrase binds Z^2 with greater affinity than Z^1, and IHF is thought to increase this binding between flanking PUs (section 1.9.5), such a BIME structure may be bound by gyrase and/or other proteins at very low concentrations.

It is not known whether motifs within a BIME act as independent units or whether they contribute additively to the overall BIME function. It would be interesting to see if there is a cooperative effect on binding of different proteins to such complex BIMEs. If not, there may not be a physiological advantage to having so many motifs within one BIME, and they may have arisen merely as a result of over-amplification. Since most BIMEs carry just one PU, and all but two have less than six (Bachellier *et al.* 1994), it may not be possible to increase superhelicity further by increasing the number of PUs. Studies of the effect of various chromosomal deletions within this BIME on chromosome organisation and *f229/gltP* gene expression may help to clarify this.

3.9.2. Induction of *f229* transcription

Transcription of *f229* was undetectable under various growth conditions, including exponential and stationary phase growth, and acid-, cold-, and osmotic-shock of non-transformed cells and cells transformed with multicopy plasmids carrying *f229*. This, along with the run-off transcription studies described in section 3.6.1, suggests that *f229* transcription is not under the control of $E\sigma^S$.

However, when *f229* was supplied on a plasmid having an internal *Stu*I deletion so as to remove all of the downstream BIME except for a single Z^2 motif (pSW15), initiation of transcription was shown to increase greatly at a plausible σ^{70}-dependent promoter sequence (fig.3.15). Enhanced transcription was also seen from a plasmid carrying an *f229*::*lacZ*α translational fusion (pSW13), which carried no BIME motifs (data not shown). It should be noted that all RNA used for primer extension was extracted from exponentially growing cells , so transcription of the majority of this RNA would have been initiated by $E\sigma^{70}$.

The above results indicate that the downstream BIME represses $E\sigma^{70}$-mediated transcription of *f229* under the conditions investigated, and that derepression occurs when most or all of the BIME motifs are removed. The most likely cause of this repression is BIME-mediated supercoiling of the promoter region, since transcription can be activated or repressed by increased levels of negative supercoiling, depending on the promoter. The $E\sigma^{70}$-initiated run-off transcripts shown in fig.3.12 were generated from linear template DNA lacking BIME motifs. The same strategy could also be adopted to investigate the effect of BIME sequences on *f229* transcription using linear template DNA, i.e. in the absence of supercoiling. In addition, plasmids carrying a strong rho-independent transcriptional terminator cloned between the *f229* promoter and various BIME motifs could be utilised to study their effects on supercoiled template DNA.

Further evidence for a role of supercoiling in the transcriptional control of *f229* came from attempted IPTG-induced transcription of *gltP* from the *lac* promoter of pSW12 (fig.3.14) which gave no detectable transcripts. Transcription from this promoter is known to be reduced by increased negative supercoiling (McClure,

1985). In contrast, IPTG-induced transcription of *fdhF* from the *lac* promoter of pSW3 was detectable and allowed the mapping of the *fdhF* terminator (fig.3.13). On this plasmid, the shortest distance between the *lac* promoter and the BIME sequence is approximately 2.6kb; on pSW12, the *lac* promoter and the BIME are separated by 580bp while the *f229* promoter and the BIME are separated by 750bp. Thus the range of repression may be limited. On the chromosome the *gltP* promoter and the BIME are separated by only 1.5kb. It is not known whether the BIME affects *gltP* transcription. In any case, if this BIME is involved in chromosome architecture, *f229* and *gltP* would be in separate chromosomal domains, and might have completely different levels of superhelicity. In addition, the interaction between RNA polymerase and the *gltP* promoter may not be supercoiling sensitive.

This leads to the question of how *f229* is transcribed on the normal chromosome carrying the BIME. Is it due to an inducible transcriptional activator, which when bound to supercoiled DNA allows efficient transcription? Alternatively, perhaps *f229* is expressed in response to conditions which cause a generalised decrease in the levels of negative supercoiling.

Increased levels of *f229* transcript following the removal of most of the BIME on a plasmid did not appear to have any effect on cell growth and viability, and *f229* mRNA is almost certainly poorly translated considering its low F_{op} value. An F_{op} value of 0.52 is less than that of *lacI* (0.62), and LacI is thought to be present in 10 copies per cell (Guoy and Gautier, 1982). These findings argue against a need to keep *f229* transcriptionally silent, suggesting that the primary role of the BIME is not to control *f229* expression. It is possible that the BIME arose at this location because *f229* is inessential, and the BIME can thus be tolerated. On the other hand, strains carrying complex BIMEs near essential genes may be selected against due to transcriptional repression of their promoters.

The results obtained may be misleading because the behaviour of *f229* on a plasmid was being observed and a different supercoiled environment may exist on the chromosome. However, lack of detectable transcription under various conditions using non-plasmid strains suggests that there may also be repression of the

chromosomal copy. The effects on *f229* and *gltP* expression produced by removal of the chromosomal BIME would be interesting to examine. However, this may not be possible if the BIME plays an important role unrelated to *f229* transcriptional control.

It would be interesting to study the effect of this BIME and others on transcription of reporter genes from promoters known to be sensitive to or unaffected by increased levels of negative supercoiling. The distance between the promoter and BIME required for activation or repression could also be studied. For example, the *cat* gene encoding chloramphenicol acetyl transferase is only 790bp, similar in length to *f229*, while *lacZ*, encoding β-galactosidase, is approximately 3kb, similar to the distance between the *lac* promoter and the BIME in pSW3.

Placing BIMEs at different distances upstream of promoters may have different effects. Further to this, Boccard and Prentki's (1993) hypothesis that RIBs/RIPs reduce the levels of positive supercoils ahead of the transcription machinery could be investigated. Introducing RIBs/RIPs between diverging reporter genes would increase the levels of negative supercoils behind RNA polymerase and may serve to inhibit transcription elongation.

3.9.3. The *f229* promoter

This promoter has good homology with the consensus Pribnow box sequence, differing only at position 5 with a C instead of an A. Point mutations at this position are known to give only mild reductions in transcription (Youderian *et al.*, 1982). Indeed, results from our lab indicate that when homology is maintained at -35, a C at position 5 may even elevate transcription. This elevation appeared more pronounced when the mutation was introduced into an "extended -10" promoter (data not shown). The Pribnow box consensus sequence is based on frequency, not promoter strength.

The -35 element of the *f229* promoter, 17bp upstream from the Pribnow box and overlapping the T-rich tail of the *fdhF* terminator, has only 2/6 homology with consensus. However, the two "conserved" Ts at positions one and two are thought

to be the most important -35 bases for efficient transcription, along with the G at position three.

A second potential promoter sequence is present which overlaps the active promoter (fig.3.8). This has TATTAT at -10 (beginning at position 121 in fig 3.8), separated by 16bp from CTGGTT, i.e. 5/6 consensus at -10 and 2/6 consensus at -35. However, primer extension revealed no transcripts originating from this promoter, which may be due to the sub-optimal spacing of the promoter elements.

Interestingly, overlapping the inverted repeat of the *fdhF* terminator is a direct repeat of $TCCTTTC(N)_7TCCTTTC$ (fig.3.13b). This is centred at -54 relative to the transcription start, and is thus a potential binding site for a class I transcriptional activator. Although DNA-binding proteins are often symmetrical dimers binding quasi-symmetrical sequences, this is not always the case. λCII is a well-studied exception (Keilty and Rosenberg, 1987). An *E.coli* ECDC database search failed to identify similar sequences upstream of promoter regions that may be activated as well as *f229*. However, not all of the above sequence may be important for activator binding, making the search more difficult. For example, if $TTTC(N)_{10}TTTC$ is crucial (separated by 10bp, i.e. one turn of the DNA helix), defining 14bp rather than 8bp would bias the search against the required sequence.

It should be noted that the run-off transcription assay giving a positive result (fig. 3.12) would not be expected to contain transcriptional activators. However, the assay uses excess purified RNA polymerase which may give positive results even in the absence of an activator. On the other hand, this assay has been used to demonstrate the requirement for transcription activators at various promoters (such as Kumar *et al.*, 1994), making the latter suggestion unlikely.

An *Eco*NI restriction site is conveniently positioned such that digestion would remove all but the last C of the direct repeat sequence. Thus, the effect of deletion on transcription of *f229* from pSW15 should be easy to study. However, even if this sequence is found to be important, that will give no further insight into the possible function of *f229* unless the same activator sequence is found in the promoter region

of an already characterised gene, or the putative activator and its gene identified and studied.

3.9.4. Transcription termination within BIME

Attempts to study transcriptional termination of *f229* using nuclease-S1 analysis produced no positive results, presumably because *f229* transcription was very weak due to the inhibitory effect of the BIME on transcription initiation. Termination of IPTG-induced transcription from the *lac* promoter of pSW11 (fig.3.14) was also undetectable, presumably because the inhibitory effect of the BIME extends to this supercoiling-sensitive promoter which is only some 50bp upstream of the *f229* promoter on pSW11 DNA. Indeed, initiation at the *lac* promoter on pSW11 was not detectable by primer extension analysis (data not shown) so negative results were not confined to the S1 mapping analyses. Likewise, termini of *gltP* transcripts initiated from the *lac* promoter on pSW12 were not detectable, perhaps for the same reason.

S1 mapping of the *gltP* transcriptional terminator using RNA from non-transformed cells gave two weak bands corresponding to sequences within the first Z^2 motif of the BIME proximal to *gltP* (although with considerable read-through) but not within the PU*-like sequence preceding it. Interestingly, the end point comes after a stem-loop and two Ts. Although PUs were not previously thought to be involved in transcription termination (except for PU*), because of their occurrences within operons, the RIB/RIP sequences which are located only at the end of transcription units may play a part in termination. Extra proteins bound to RIBs/RIPs, compared to other BIMEs, may facilitate this. Possible Rho-dependent termination sites have been postulated within BIMEs (Gilson *et al.,* 1986), but direct experimental data have been unavailable.

If transcription does extend into RIBs/RIPs, there may be a physiological reason for having BIME motif sequences within the mRNA. Perhaps proteins that can specifically bind to the PU and L RNA sequences confer a special property on such transcripts.

3.9.5 Translation of *f229*

Minicell experiments detected no protein of the size expected for F229, except when using strains carrying plasmids known to give elevated levels of *f229* transcript. Expression of F229 was detectable in a T7 promoter-based system (using rifampicin to suppress normal transcription) and in an *in vitro* transcription/ translation system. Moreover, an *f229::lacZα* gene fusion protein was able to give α-complementation. Thus there is initiation of *f229* translation, albeit at very low levels. Under most growth conditions, translation of the low levels of *f229* transcript present is expected to be limited, considering its poor F_{op}. The rare tRNA species required to translate the minor codons of *f229* would probably be sequestered by more abundant transcripts. It may be that F229 can perform its function at low concentrations, as is the case for example with LacI.

On the other hand, it is possible that F229 protein levels could increase under conditions where specific translation of *f229* mRNA is favoured. This situation occurs in the expression of *cspA* which encodes the major cold-shock protein. Following temperature downshift, the majority of protein synthesis ceases. Ribosomes appear to be modified so as to preferentially translate *cspA* mRNA, although it is not yet known if translation of other cold-shock transcripts is also enhanced (section 1.11.1). The *f229* transcript contains a Y-box sequence present in certain cold-shock gene promoter regions, although *f229* transcription is not cold-shock-inducible. In principle, enhanced translation of low mRNA levels by modified ribosomes might give a considerable increase in F229 protein production. However, minicells carrying pSW15 did not yield increased amounts of F229 when labelled at various times following cold-shock. On the other hand if other factors are required, such as a cold-shock-inducible protein to transduce the signal, a negative result might be expected in minicells since they contain no chromosomal DNA to encode such a protein. In addition, CspA binds Y-box sequences and has been proposed as an RNA chaperone, which may assist in the unwinding of cold-stabilised mRNA structures. Thus the increased translation of cold-shock transcripts at reduced temperature may be reliant upon functions not present in minicells. In this respect,

S30 extracts from cold-shocked cells may be a better system to test for enhanced *f229* translation following cold-shock. Another possibility would be to label proteins encoded by T7 RNA polymerase-generated *f229* transcripts at low temperature, although the unnaturally high levels of *f229* mRNA might mask enhancement of translation.

3.9.6 Function of F229

Various computer searches failed to reveal amino acid sequence homology with other proteins, or with conserved protein domains. However, Blattner *et al.* (1993) noticed that F229 has a potential N-terminal signal-peptide-like sequence if translation initiates at the first methionine in the open reading frame (fig. 3.8). This would give an N-terminal sequence including two positively charged lysine residues able to interact with the negatively charged phosphate groups at the surface of the inner membrane, followed by twelve hydrophobic residues able to initiate insertion into the membrane (Michaelis and Beckwith, 1982). However, initiation at the downstream leucine codon (section 3.4) would leave just four of the hydrophobic residues thus presumably eliminating this potential signal sequence. Such a protein would have a molecular weight of 23.8kDa, which cannot be distinguished from a protein of 25.1kDa by the SDS-PAGE analysis presented in section 3.7. Determination of the N-terminal sequence of F229 might therefore be informative regarding the likely localisation of this protein within the cell, and give some indication as to its possible function, although the low protein output would make such work difficult. Immunological approaches might help to overcome this. For example, specific anti-β-galactosidase monoclonal antibodies covalently bound to a column might be used to purify an F229::LacZ fusion protein, which could then be subjected to N-terminal sequence analysis. Alternatively, antibodies raised against a synthetic peptide comprising the first 10 amino acids of the putative F229 signal sequence could be used to probe a Western blot of F229 protein. Antibodies specifically cross-reacting with the ~25kDa protein would provide evidence that the

protein does indeed have an N-terminal signal-sequence, whilst negative results would argue against (although not disprove) the presence of this sequence.

Lack of a consensus Shine-Dalgarno ribosome binding site argues against initiation at the first methionine. It may be possible to modify the *in vitro* transcription/ translation system described in section 2.10.3 by hybridising newly synthesised transcripts with oligonucleotides before proceeding to the translation reaction. Oligonucleotides directed against the two putative ribosome-binding sequences within the mRNA would be expected to interfere with translation, thus reducing protein output. While an oligonucleotide bound at the putative downstream ribosome binding sequence might also affect translation initiated from the upstream putative sequence (by preventing ribosomes from continuing along the transcript), an oligonucleotide bound upstream would be less likely to interfere with translation initiated further downstream.

Other methods could be employed to locate F229 within the cell. For example, the transposon Tn*phoA* can be used to specifically detect bacterial genes that code for cell envelope proteins (Manoil and Beckwith, 1985). Tn*phoA* inserts into a chromosomal gene and, when in the correct frame, will fuse alkaline phosphatase to the N-terminus of the protein product of that gene. Alkaline phosphatase is enzymically active only if fused to a sequence that promotes export of the protein into the envelope of the cell (Manoil and Beckwith, 1986). An *f229::phoA* fusion carrying only the putative signal sequence of *f229* should not suffer from the potential translation problems arising from rare codons in *f229*..

Certain stimuli may lead to increased *f229* expression. These would have to generate mechanisms first of all to overcome BIME-mediated transcription repression, and then to compensate for the normally poor translatability of the *f229* mRNA. Identification of conditions that induce *f229* expression would provide further insight into the fuction of this gene, but lack of amino acid sequence homology with characterised proteins makes it difficult to choose a logical starting point. It seems more likely that F229 is non-inducible and is not required in

elevated amounts, which would make the characterisation of its function all the more difficult.

3.10. The λ440 probe-positive region

Fig. 3.1. shows that there is strong hybridisation between the sigma probe 746J and λ440 DNA, although not with λ439 or λ441, indicating that the probe-positive region lies in the insert DNA unique to λ440. Southern analysis in this lab by David Smillie fine-mapped the positive region to a 2.8kb *Pst*I fragment (data not shown). However, this and other positive fragments were found to be unclonable, and so the λ440 probe-positive region was not investigated further.

More recently the genomic DNA sequence of *E.coli* K12 strain MG1655 from 58.3 to 59.0 minutes was published as part of the *E.coli* genome project (EMBL accession number U36840). This region spans the entire λ440 insert DNA from strain W3110. However, certain discrepancies were found between the two strains, most notably an ~6kb deletion in MG1655.

By comparing the restriction maps of the published MG1655 sequence and of W3110, the present author mapped the deletion between two *Pst*I sites at 2771.9kb and 2779.3kb on the physical map of W3110 (Rudd, 1991). This deletion completely removes the 2.8kb probe-positive region. It is possible that a gene encoding a toxic protein resides in this region, hence the failure to obtain subclones from λ440; conceivably the "deletion" in MG1655 is also related to this, and may have arisen in the process of sequencing. Although this gene is not necessarily the one hybridising with 746J, the observed deletion in MG1655 (if genuine) suggests that the probe-positive region is inessential in this strain.

The *E.coli* genome project intends to sequence the corresponding region in W3110, and inspection of the deleted section may reveal a candidate sigma gene as well as the potentially toxic gene. In any case, none of the open reading frames within the published MG1655 sequence that are also present in the λ440 insert DNA share any significant amino acid sequence homology with bacterial sigma factors (data not shown).

Chapter 4

4.1. Introduction

As discussed in sections 3.1 and 3.2.1, oligonucleotides directed against the conserved core-binding subregion 2.1 of bacterial sigma factors were used to probe the library of Kohara *et al.* (1987) in an attempt to locate novel sigma genes of *E.coli.* In this chapter I describe the mapping, sequencing and characterisation of a candidate sigma gene, *orf243*. Part of this work was carried out by a summer student, Thomas Pratt. Work by him is acknowledged in the appropriate sections.

4.2.1. Design of oligonucleotide probes

As discussed in section 3.2.1, the second oligonucleotide probe, 745J, carried the most optimal codon for each amino acid within the subregion 2.1 target sequence. The sequence of 745J is:

5'- GGT GAC CAG GAA GCT CGT AAA ACC CTG ATC GAA GCT AAC CTG CGT CTG GTT GTT TCT ATC GCT AAA AAA TAC GCT AAC-3'

The bases which differ from 746J are underlined.

4.2.2. Probing *E.coli* genomic DNA

745J was 5′ end-labelled using [γ-^{32}P] ATP as described in section 2.8.2 and hybridised with nylon filters carrying DNA from the miniset clones of the *E.coli* K12 genome (Kohara *et al.*, 1987). This probe hybridised with three pairs of contiguous clones: λ139/λ140, λ433/λ434, and λ509/510 (fig. 4.1). The overlapping insert DNA of λ509/510 is centred near 3375kb on the physical map (67.2 minutes on the genetic map) and carries *rpoD*, the gene encoding σ^{70}. The nucleotide sequence alignment between *rpoD* and 745J is shown in fig. 4.2.

The overlapping insert DNA of λ433/λ434 is centred near 2700kb on the physical map (55.3 minutes on the genetic map) and carries *purL*, encoding formylglycineamide ribonucleotide synthetase. This enzyme is involved in purine biosynthesis and is not a sigma factor. However, *purL* has significant sequence homology with 745J as determined by Fasta and BLAST nucleotide sequence alignments, scoring higher than several bacterial sigma factors (data not shown).

The overlapping insert DNA of λ139/λ140 is centred near 384kb on the physical map (8.2 minutes on the genetic map). The location of a candidate sigma gene in this unsequenced region was investigated further.

4.2.3. Mapping and subcloning of the λ139/λ140 probe-positive region

The probe-binding region was mapped more accurately by Thomas Pratt using Southern analysis as described in section 2.8.3. Using "Kohara" restriction enzymes, 745J was shown to hybridise with a 3.3kb *Eco*RV/*Pvu*II fragment (data not shown). Fig. 4.3 shows the λ139/λ140 insert region with the probe-positive fragment highlighted. The larger 4.2kb *Eco*RV fragment was subcloned into the *Eco*RV site of pBluescript SK(-) by Thomas Pratt to generate pTEP4.

Figure 4.1. Location of a candidate sigma gene within the miniset clones of the *E.coli* genome. Hybridisation was performed at 50°C for 18 hours. Final washing conditions were 0.1x SSC, 0.1% SDS, for 2 minutes at 65°C. The positions of λ clones 139, 140, 433, 434, 509, and 510 are marked. The overlapping insert DNA of λ433/λ434 carries *purL*; the overlapping insert DNA of λ509/λ510 carries *rpoD*.

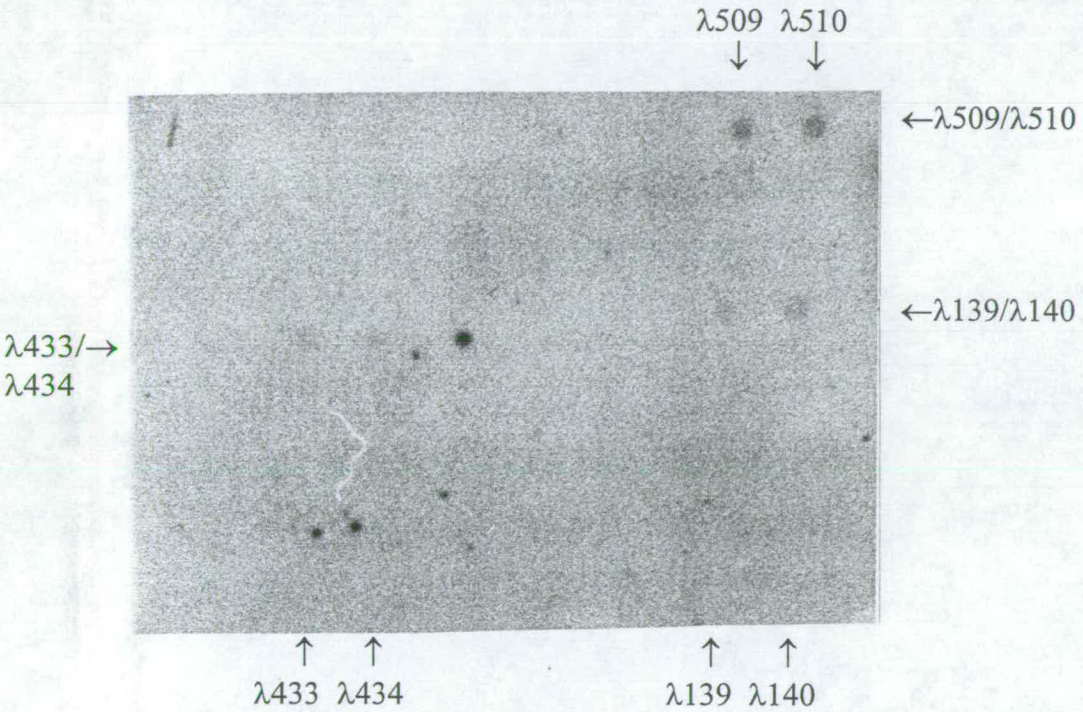
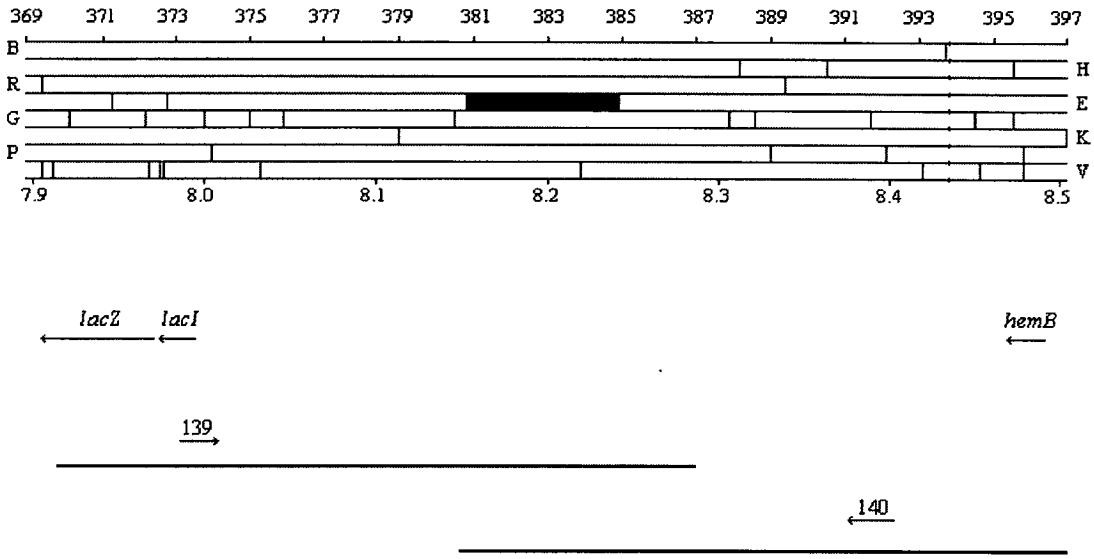


Figure 4.2. Nucleotide sequence alignment (5′→3′) between *rpoD* (top line) and 745J. The *rpoD* sequence shown begins at nucleotide number 1080 of the EMBL sequence, accession number U23083. The 745J sequence begins at nucleotide number 1. Identical bases are shown in bold.

1080-G AAA**GCCCGCCGTGCGCAGAAAGAGATGGTTGAAGCGAAC**
GGT**GACCAGGAAGCTCGTAAA** ACCC**TGATCGAAGCTAAC**

TTACGTCTGGTTATTTCTATCGCTAAGAAATACACCAAC CG-1160
CTGCGTCTGGTTGTTTCTATCGCTAAAAAATACGCTAAC

Figure 4.3. Restriction map of *E.coli* chromosomal inserts cloned in λ139 and λ140, based on the map of Rudd (1991). The top scale shows the coordinates in kb on the physical map of the *E.coli* K12 chromosome; the bottom scale shows the coordinates in minutes on the genetic map. The *Eco*RV fragment subcloned into pBluescript SK(-) to give pTEP4 is highlighted, and the length and orientation of the chromosomal inserts carried by λ139 and λ140 are indicated below the map. The following abbreviations are used for restriction enzymes: B, *Bam*HI; E, *Eco*RV; H, *Hind*III; G, *Bgl*II; K, *Kpn*I; P, *Pst*I; R, EcoRI; V, *Pvu*II.



4.2.4 Fine mapping of the probe-positive region

To determine further the location of the probe-positive region, the author digested pTEP4 with "non-Kohara" enzymes to identify other restriction sites (fig. 4.4). Fig. 4.5 shows the plasmid map of pTEP4 as determined from this restriction analysis. After identifying convenient restriction sites within the insert DNA, pTEP4 was subjected to Southern hybridisation analysis using 745J as probe (fig. 4.6a). Additional Southern analysis located the probe-positive region on a 350bp *Xho*I/*Stu*I fragment (fig. 4.6b).

Figure 4.4. Restriction analysis of pTEP4 DNA by 0.6% agarose gel electrophoresis. Restriction enzymes used are as follows:
Track 1, *Bgl*I; Track 3, *Eco*RV+*Stu*I; Track 4, *Pvu*II; Track 5, *Eco*RV+*Pvu*II; Track 6, *Pst*I; Track 8, *Bgl*II+*Eco*RV; Track 9, *Xho*I; Track 10, *Pst*I+*Xho*I; Track 11, *Bam*HI+*Kpn*I; Track 13, *Kpn*I. Tracks 2, 7, and 12 contain 1kb marker DNA (Gibco BRL). Marker fragment sizes are given in kilobase pairs.

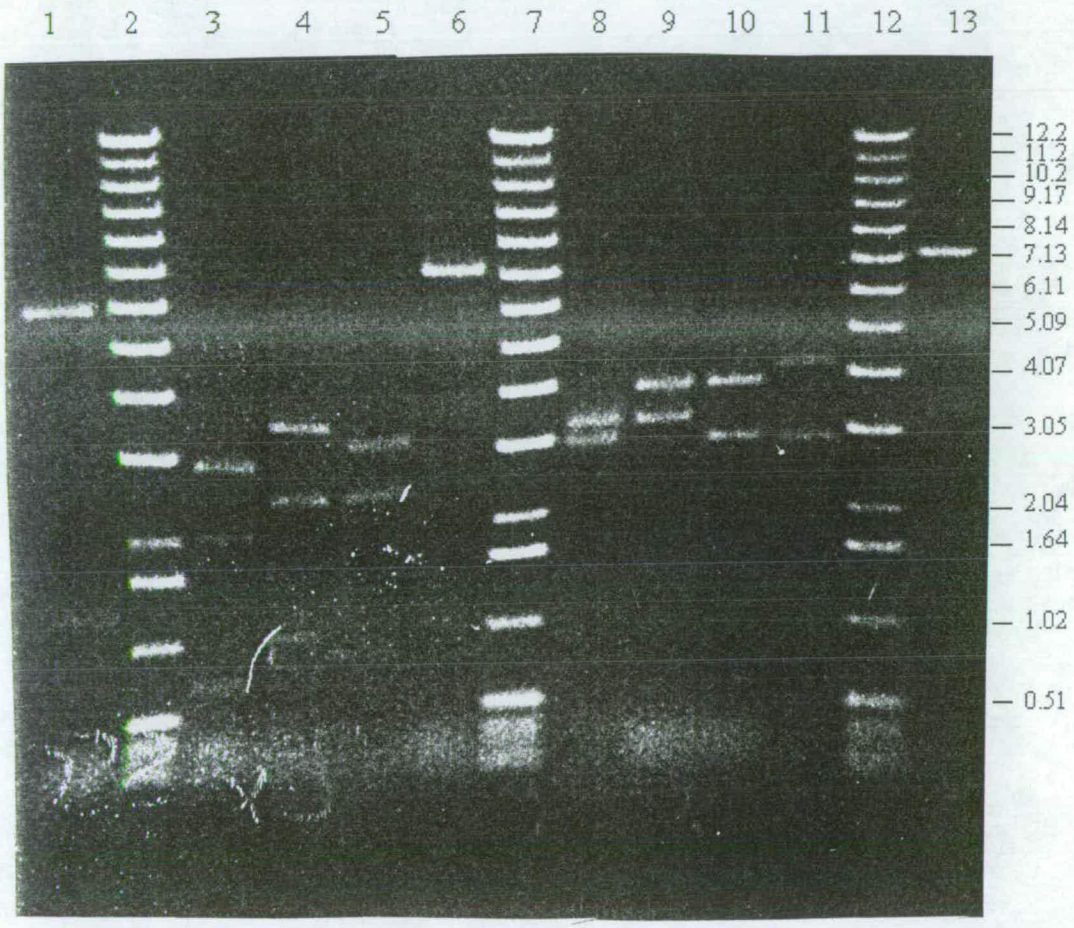
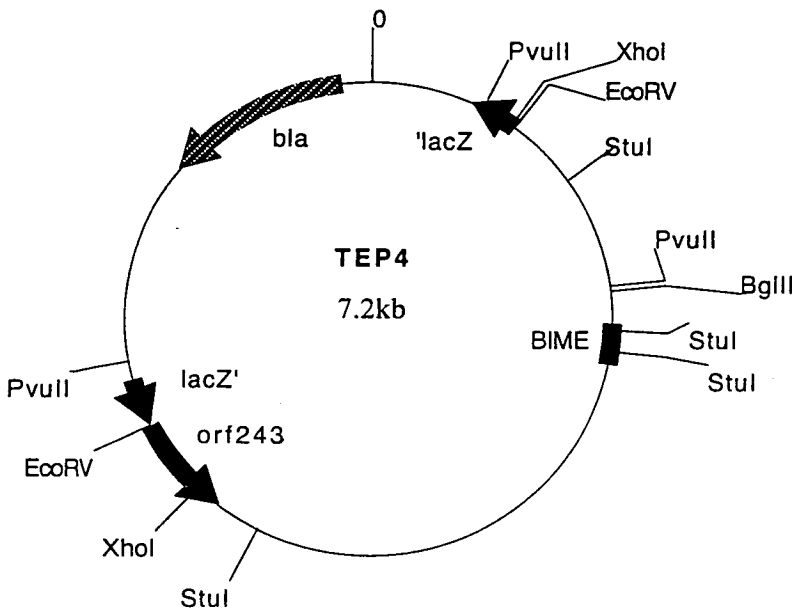


Figure 4.5. Plasmid map of pTEP4 as determined by restriction analysis (fig. 4.4). The position and orientation of *orf243* (see text) are marked.
Abbreviations for restriction enzymes are as follows:
B, *Bgl*II; E, *Eco*RV; P, *Pvu*II; S, *Stu*I; X, *Xho*I.

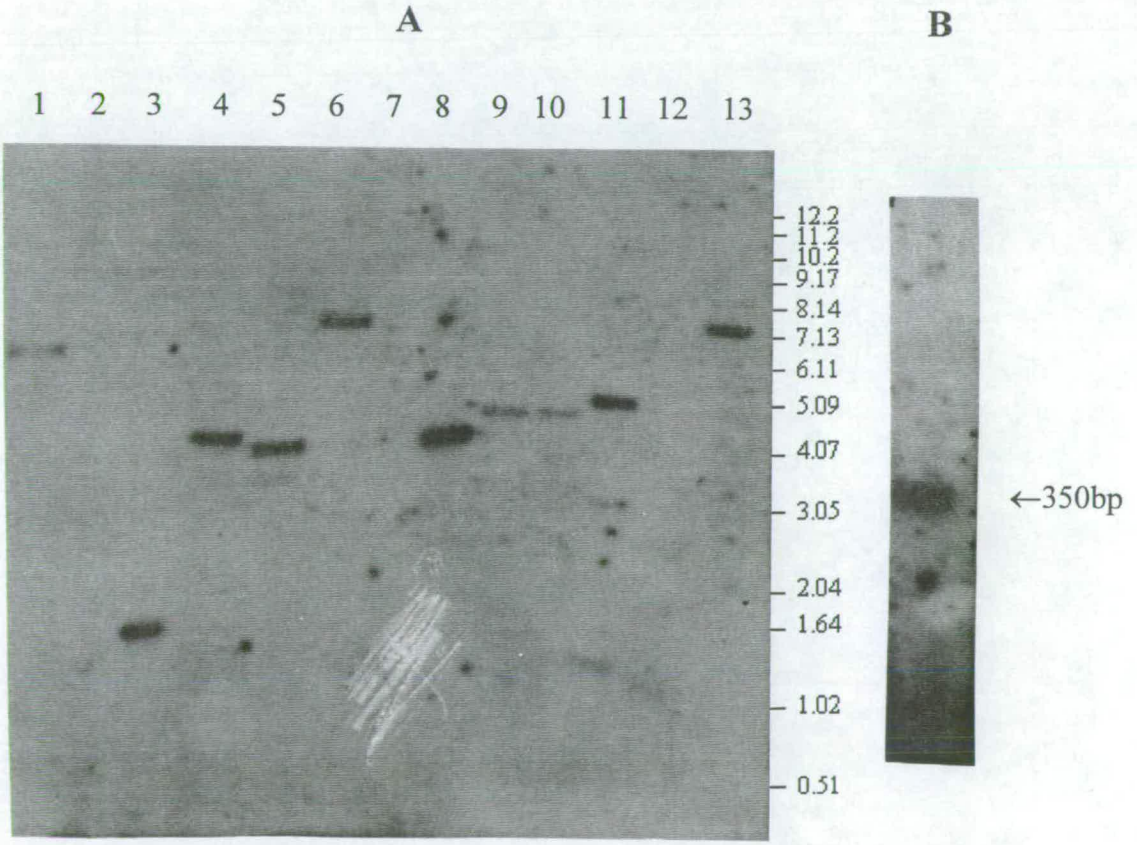


Restriction sites (bp) :

*Pvu*II	530, 1580, 5000.	*Stu*I	1050, 1800, 1900, 4000.
*Xho*I	670, 4330.	*Bgl*II	1610.
*Eco*RV	700, 4700.		

Figure 4.6. Fine mapping of the 745J probe-positive region within pTEP4.
a. Southern hybridisation of pTEP4 DNA. DNA was transferred from the gel shown in fig. 4.4 to a nitrocellulose filter as described in section 2.8.3. Hybridisation with [^{32}P]-labelled 745J was performed at 50°C for 18 hours. Final washing conditions were 0.1x SSC, 0.1% SDS, for 2 minutes at 65°C. Restriction enzymes used are as follows:
Track 1, *Bgl*I; Track 3, *Eco*RV+*Stu*I; Track 4, *Pvu*II; Track 5, *Eco*RV+*Pvu*II; Track 6, *Pst*I; Track 8, *Bgl*II+*Eco*RV; Track 9, *Xho*I; Track 10, *Pst*I+*Xho*I; Track 11, *Bam*HI+*Kpn*I; Track 13, *Kpn*I. Tracks 2, 7, and 12 contain 1kb marker DNA (Gibco BRL).

b. To further map the probe-positive region pTEP4 DNA was digested with *Xho*I+*Stu*I and subjected to 0.6% agarose gel electrophoresis (data not shown). DNA was transferred to a nitrocellulose filter as described in section 2.8.3. Hybridisation and washing were performed as above. The 350bp probe-positive fragment is indicated.



4.3. Nucleotide sequence of the probe-positive region

Double stranded DNA sequencing of pTEP4 was initiated using reverse primer as described in section 2.7. The original strategy was to sequence out of the vector towards the probe-positive region, at least 250bp away. However, when the first sequence data (not shown) were subjected to computer analysis, it was found that the entire λ140 insert DNA sequence had just been submitted (GenEMBL accession number D64043). GCG Fasta and Bestfit programs were employed to locate the 745J binding site within the previously identified *Xho*I/*Stu*I fragment. This was located just downstream from the *Xho*I site (fig. 4.7). As with 746J/*f229* interaction (section 3.5), it appears that homology towards the middle of the probe is of primary importance with additional contacts being made along the entire length of the probe.

Figure 4.7. Nucleotide sequence alignment between *orf243* (top line) and 745J. The *orf243* sequence shown begins at the *Xho*I target site at position 582 in the GenEMBL sequence, accession number D64043 (also in fig. 4.8). The 745J sequence begins at position 29 (section 4.2.1). Perfect matches are shown in bold.

*Xho*I
↓
582-CTCG**AGCT**(-)T**CCTGCGTC**

29-TCGA**AGCT**AA**CCTGCGTC**

4.4. 745J hybridises to an open reading frame

The probe-positive region was subjected to further sequence analysis. Although previously unidentified, an open reading frame was found with the potential to encode 243 amino acids, and called *orf243* (fig. 4.8). The upstream portion of this open reading frame would lie in the unsequenced region of λ139; lack of promoter and ribosome-binding sequences may explain why it was not noticed before.

A BLAST search with the amino acid sequence of *orf 243* revealed striking identity (80% range) with a family of *Pseudomonas* enzymes, the 4-hydroxy-2-oxovalerate aldolases. The sequence alignment between Orf243 and these enzymes, as determined by the GCG Pileup program, is shown in fig. 4.9. No other open reading frames were found in the ~3kb of DNA downstream of *orf243* that is present in pTEP4, either by the author or by the group who sequenced the λ140 insert DNA. However, a BIME was found approximately 2.2kb downstream of *orf243* (see fig. 4.5). This has the structure $[Z^2SY]_2$ and is a member of the BIME-2 family (section 1.9.3).

4.5. Proteins encoded by pTEP4

Proteins labelled in minicells (section 2.10.1) carrying pTEP4 included no detectable proteins that were not labelled in minicells carrying pBluescript SK (data not shown). This is unsurprising for *orf243*, since promoter and ribosome-binding sites are not likely to be present. It also agrees with the apparent absence of other open reading frames in the pTEP4 insert DNA. The plasmid map of pTEP4 given in fig. 4.5 shows the *lac* promoter of pBluescript in the same orientation as *orf243*. However, IPTG had no effect on protein output from minicells transformed with pTEP4. No obvious transcriptional terminators were found downstream of *orf243*. This further indicates that there are no open reading frames in the same orientation as *orf 243*, which might have been missed in the absence of IPTG through weakness of their normal promoters.

Figure 4.8. Nucleotide sequence of *orf243* from the GenEMBL entry, accession number D64043. The amino acid sequence encoded by the open reading frame is given below the relevant codons. Target sites for *Eco*RV and *Xho*I are indicated.

```
          10                  30                  50
           .   .   .   .   .   .   .   .   .   .
GATCTGAAAAATGCCTGGCAGGCTGGCGCGCGGGTGGTTCGTGTGGCAAC
D  L  K  N  A  W  Q  A  G  A  R  V  V  R  V  A  T

                    70                  90
                .   .   .   .   .   .   .   .   .
GCACTGTACCGAAGCTGATGTTTCCGCCCAGCATATTCAGTATGCCCGCG
 H  C  T  E  A  D  V  S  A  Q  H  I  Q  Y  A  R  E

          110                 130                 150
           .   .   .   .   .   .   .   .   .   .
AGCTCGGAATGGACACCGTTGGTTTTCTGATGATGAGCCATATGACCACG
  L  G  M  D  T  V  G  F  L  M  M  S  H  M  T  T

                    170                 190
                .   .   .   .   .   .   .   .   .
CCGGAGAATCTCGCCAAGCAGGCAAAGCTGATGGAAGGCTACGGTGCGAC
P  E  N  L  A  K  Q  A  K  L  M  E  G  Y  G  A  T

          210                 230           EcoRV
           .   .   .   .   .   .   .   .   ↓   .
CTGTATTTATGTGGTGGATTCTGGCGGTGCGATGAACATGAGCGATATCC
 C  I  Y  V  V  D  S  G  G  A  M  N  M  S  D  I  R

                    270                 290
                .   .   .   .   .   .   .   .   .
GTGACCGTTTCCGCGCCCTGAAAGCAGAGCTGAAACCAGAAACGCAAACT
  D  R  F  R  A  L  K  A  E  L  K  P  E  T  Q  T

          310                 330                 350
           .   .   .   .   .   .   .   .   .   .
GGCATGCACGCTCACCATAACCTGAGTCTTGGCGTGGCGAACTCTATCGC
G  M  H  A  H  H  N  L  S  L  G  V  A  N  S  I  A
```

```
                                                                    .
GGCGGTGGAAGAGGGCTGCGACCGAATCGACGCCAGCCTCGCGGGAATGG
 A  V  E  E  G  C  D  R  I  D  A  S  L  A  G  M  G

       410                   430                   450
        .                     .                     .
GCGCGGGCGCAGGTAACGCACCGCTGGAAGTGTTTATTGCCGCCGCGGAT
  A  G  A  G  N  A  P  L  E  V  F  I  A  A  A  D

                470                   490
                 .                     .                     .
AAACTGGGCTGGCAGCATGGGACCGATCTCTATGCGTTAATGGATGCCGC
K  L  G  W  Q  H  G  T  D  L  Y  A  L  M  D  A  A

       510                   530                   550
        .                     .                     .
CGACGACCTGGTGCGTCCGTTGCAGGATCGACCGGTACGAGTCGATCGCG
 D  D  L  V  R  P  L  Q  D  R  P  V  R  V  D  R  E

                570                   XhoI  590
                 .                     .  ↓  .                     .
AAACGCTGGCGCTGGGATACGCTGGTGTTTACTCGAGCTTCCTGCGTCAC
  T  L  A  L  G  Y  A  G  V  Y  S  S  F  L  R  H

       610                   630                   650
        .                     .                     .
TGTGAAACGGCGGCGGCGCGTTATGGCTTAAGTGCGGTGGATATTCTCGT
C  E  T  A  A  A  R  Y  G  L  S  A  V  D  I  L  V

                670                   690
                 .                     .                     .
TGAGCTGGGCAAACGCCGGATGGTTGGCGGCCAGGAGGATATGATCGTTG
 E  L  G  K  R  R  M  V  G  G  Q  E  D  M  I  V  D

       710                   730
        .                     .
ACGTGGCGCTGGATCTGCGCAACAACAAATAA
 V  A  L  D  L  R  N  N  K  *
```

Figure 4.9. Alignment of 4-hydroxy-2-oxovalerate aldolase protein sequences from *Pseudomonas* and the putative product of the *E.coli* open reading frame, *orf 243*, using GCG PILEUP. The *Pseudomonas* enzymes are involved in the degradation of the following substrates : BphI, biphenyls and polychlorinated biphenyls; NahM, naphthalene; DmpG, phenols; XylK, xylenes; TodH, toluenes; CmtG, pCumate. See section 1.12 for further details.

```
  nahm   .MNLHGKSVI LHDMSLRDGM HAKRHQISLE QMVAVATGLD QAGMPLIEIT
  dmpg   MTFNPSKKLY ISDVTLRDGS HAIRHQYTLD DVRAIARALD KAKVDSIEVA
  xylk   MTFNPGKKLY ISDVTLRDGS HAIRHQYSIQ NVQDIARALD KARVDSIEVT
Orf243   .......... .......... .......... .......... ..........
  todh   ...MTQQKLY ISDVTLRDGS HAIRHQYTVE QVKQIARALD DAKVDSIEVA
  cmtg   .MFDTRKKIY VSDVTLRDGM HAVRHQYSLA DAERIARALD EAGVDSIEVA

         51                                                  100
  bphi   HGDGLGGSSV NYGFPAHSDE EYLGAVIPLM KQAKVSALLL PGIGTVEHLK
  nahm   HGDGLGGRSI NYGFPAHSDE EYLRAVIPQL KQAKVSALLL PGIGTVDHLK
  dmpg   HGDGLQGSSF NYGFGRHTDL EYIEAVAGEI SHAQIATLLL PGIGSVHDLK
  xylk   HGDGLQGSSF NYGFGAHSDL EWIEAAADVI QHARVTVLLV PGIGTVHDLK
Orf243   .......... .......... .......... .......... .......DLK
  todh   HGDGLQGGSF NYGFGAQSDL EWIEAAASVV KASKIATLLL PGIGTVHDLK
  cmtg   HGDGLQGSSF NYGFGAHTDL EWIERVAATV RRAKIATLLL PGIGTVHDLK

         101                                                 150
  bphi   MAKDLGVNTI RVATHCTEAD VSEQHITQSR KLGLDTVGFL MMAHMASPEK
  nahm   MALDCGVSTI RVATHCTEAD VSEQHIGMAR KLGVDTVGFL MMAHMISAEK
  dmpg   NAYQAGARVV RVATHCTEAD VSKQHIEYAR NLGMDTVGFL MMSHMIPAEK
  xylk   AAYDAGARSV RVATHCTEAD VSRQHIEYAR ELGMDTVGFL MMSHMIPAEQ
Orf243   NAWQAGARVV RVATHCTEAD VSAQHIQYAR ELGMDTVGFL MMSHMTTPEN
  todh   AVYEAGVRVV RVATHCTEAD ISRQHIEYAR HLGMEAVGFL MMSHMTTPQH
  cmtg   NANAAGASVV RVATHCTEAD ISQQHIEYAR KLGMDTVGFL MMSHMTTPTA

         151                                                 200
  bphi   LVSQALLMQG YGANCIYVTD SAGYMLPDDV KARLSAVRAA LKPETELGFH
  nahm   VLEQAKLMES YGANCIYCTD SAGYMLPDEV SEKIGLLRAE LNPATEVGFH
  dmpg   LAEQGKLMES YGATCIYMAD SGGAMSMNDI RDRMRAFKAV LKPETQVGMH
  xylk   LAAQGKLMET YGAQCIYMAD SGGAMNMNDI RDRMRAFKAV LNPQTQTGMH
Orf243   LAKQAKLMEG YGATCIYVVD SGGAMNMSDI RDRFRALKAE LKPETQTGMH
  todh   LAQQAKLMES YGATVCYVVD SGGALSMNDV RDRFRAFKDV LKPETQTGMH
  cmtg   LAVEAKKMES YGAQCIYVVD SGGAMNMYDI ADRFKALKDV LDPSTQTGMH

         201                                                 250
  bphi   GHHNLAMGVA NSIAAIEAGA TRIDAAAAGL GAGAGNTPME VFIAVCARMG
  nahm   GHHNMGMAIA NSLAAIEAGA ARIDGSVAGL GAGAGNTPLE VFVAVCKRMG
  dmpg   AHHNLSLGVA NSIVAVEEGC DRVDASLAGM GAGAGNAPLE VFIAVAERLG
  xylk   AHHNLSLGVA NSIIAVEEGC DRIDASLAGM GAGAGNAPLE VFIAAAERLG
Orf243   AHHNLSLGVA NSIAAVEEGC DRIDASLAGM GAGAGNAPLE VFIAAADKLG
  todh   AHHNLSLGVA NSIVAVENGC DRVDASLAGM GAGAGNAPLE VFIAAAERMG
  cmtg   AHHNLSLGVA NSIVALEYGC DRIDASLTGM GAGAGNAPLE VFIAAVDRMG

         251                                                 300
  bphi   IETGVDVFKI QDVSEDLVVP IMDHVIRIDR DSLTLGYAGV YSSFLLFAKR
  nahm   VETGIDLYKI MDVAEDLVVP MMDQPIRVDR DALTLGYAGV YSSFLLFAQR
  dmpg   WNHGTDLYTL MDAADDIVRP LQDRPVRVDR ETLGLGYAGV YSSFLRHAEI
  xylk   WNHGTDLYTL MDAADDIVRP LQDRPVRVDR ETLGLGYAGV YSSFLRHAEV
Orf243   WQHGTDLYAL MDAADDLVRP LQDRPVRVDR ETLALGYAGV YSSFLRHCET
  todh   WNHGTDLYKL MDAADDLVRP LQDRPVRVDR ETLALGYAGV YSSFLRHSEM
  cmtg   LKHGCDVRKL IDAAEEIVRP LQERPVRVDR ETLALGYAGV YSSFLRHTEA

         301                                                 350
  bphi   ASEKYGVPAR DILVELGRRG MVGGQEDMIE DTAMTMARER GLTLTAA*..
  nahm   AEKKYGVSAR DILVELGRRG TVGGQEDMIE DLALDMARAR QQQKVSA*..
  dmpg   AAAKYNLKTL DILVELGHRR MVGGQEDMIV DVALDLLAAH KENRA*....
  xylk   AAAKYGLKTL DILVELGRRR MVGGQEDMIV DVALDLLAAR KEQQA*....
Orf243   AAARYGLSAV DILVELGKRR MVGGQEDMIV DVALDLRNNK *.........
  todh   AASKYGLKTV DILVELGRRR MVGGQEDMII DVALDLLKQQ EHEGIRSEPV
  cmtg   AAHKYGLDAF EILVELGRRR MVGGQEDMIV DVALDMMSRK PQGTMRDVIS

         351
  bphi   ......
  nahm   ......
  dmpg   ......
  xylk   ......
Orf243   ......
  todh   STEAN*
```

4.6 Discussion

The major codon probe designed to locate novel sigma genes of *E.coli* was found to hybridize strongly to the overlap region of λ139/λ140. The probe-binding site was fine-mapped and sequenced, and was located just downstream of an *Xho*I site in the GenEMBL sequence D64043 (fig. 4.7). This region at 8.13 minutes on the physical map carries a previously unidentified open reading frame potentially encoding at least 243 amino acids, although the upstream sequence is presently unknown. A BLAST search of various databases to identify similar sequences revealed 84% identity with the *Pseudomonas putida dmpG* gene product (fig.4.9). *dmpG* is carried on plasmid pVI150 of strain CF600, and encodes the 4-hydroxy-2-oxovalerate aldolase enzyme involved in the *meta*-cleavage pathway of phenol and its methyl derivatives (Shingler *et al.,* 1992; fig. 1.5). The protein product potentially encoded by the open reading frame, Orf243, also has similar identity with the isofunctional *meta*-cleavage enzymes XylK, TodH, BphF, and BphI (section 1.12 and fig. 4.9). The sequence identities are known to be in the 80% range, indicating striking conservation (Lau *et al.,* 1994).

If the homology extends throughout the entire length of the two proteins, Orf243 is predicted to have a further 97 N-terminal amino acids, which would be encoded by DNA in the unsequenced region upstream of the λ140 insert. Since all of the *meta*-cleavage genes studied so far are encoded by a single operon, it is possible that other genes of this putative *meta*-cleavage operon are also present in this 6.5kb region lying between *lacI* and the *orf243* coding sequence. No open reading frames having sequence similarity to either *meta*-cleavage or other proteins have been found in ~3kb downstream of *orf243*. Likewise, no detectable levels of protein were found in minicells carrying pTEP4. Comparisons of various *meta*-cleavage operons of *Pseudomonas* show that the order of the genes does not always directly correspond to the order of the pathway reactions that their products catalyse. Additionally, the actual order of analogous genes is not conserved between different operons. In their respective operons, *todH* is last, *bphI* is second last and *dmpG* and *xylK* are third last. Perhaps in this putative *E.coli* operon, *orf243* lies at the downstream end.

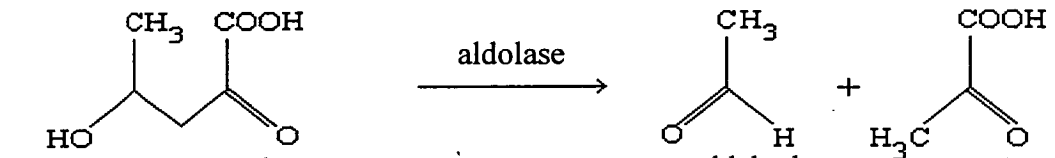
Within the known pathways, the acylating aldehyde dehydrogenase (AADH) enzymes catalysing the conversion of acetaldehyde to acetyl-CoA are also highly conserved. Lau *et al.* (1994) showed that TodI shares 75% amino acid identity with both DmpF and XylQ, and 85% identity with BphG. There is no significantly similar Orf in the known genomic sequence of *E.coli* K12. The nucleotide sequence of the region upstream of *orf243* should therefore be determined and examined for possible homologues of AADH and of other known *meta*-cleavage enzymes.

If it appears that *orf243* is part of a *meta*-cleavage operon, comparison of open reading frame amino acid sequences with enzymes of known function in other pathways may also indicate potential substrates. *E.coli* K12 is known to grow on compounds such as 3-phenylpropionic and 3-(3-hydroxyphenyl)propionic acids via a catechol intermediate (Burlingame and Chapman, 1983), and necessary enzymes including 4-hydroxy-2-oxovalerate aldolase have been identified, although the genes involved remain unmapped. Clearly *orf243* may encode the aldolase in question. The ability of K12 to grow on these compounds and others could be investigated in strains with chromosomal deletions of *orf243*. The induction of transcription by various substrates could also be investigated. In addition, the protein output of subclones carrying the upstream region expected to contain the remaining genes of the putative *meta*-cleavage operon could be analysed in a T7 expression system. The molecular weights and other characteristics of any proteins present could then be compared with those of the *meta*-cleavage enzymes in *Pseudomonas*.

Although *E.coli* C can metabolize homoprotocatechuate thanks to the *hpc* operon, K12 is apparently devoid of these genes (Jenkins and Cooper, 1988). Stringfellow *et al.* (1995) found poor identity (16.2-22.3%) between the aldolase of this pathway, HpcH, and the aldolases of the *Pseudomonas* catechol pathways. Likewise, *orf243* has only 14% identity with HpcH, in contrast to its 84% identity with the catechol metabolism enzyme DmpG. The lack of identity between HpcH and the catechol pathway aldolases may be due to the structural differences between their substrates (see below). HpcH was found to have significant identity (47.8%) with the amino acid sequence of an as yet uncharacterised open reading frame in K12 known as *f256*

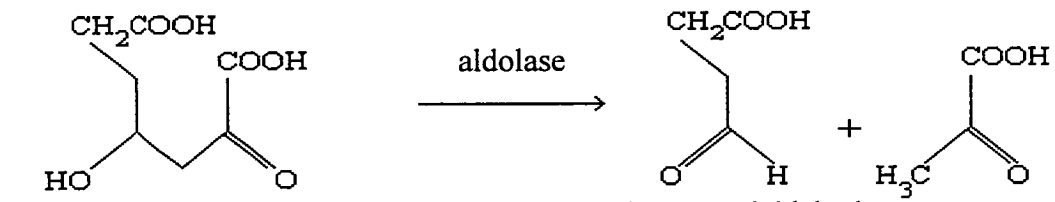
(Stringfellow *et al.*, 1995). This may encode an aldolase with substrate specificity closer to that of HpcH than DmpG. A candidate enzyme is 2-keto-3-deoxyglucarate aldolase which produces pyruvate and tartronic semialdehyde during the catabolism of glucarate (Blumenthal and Fish, 1963). *f256* maps to 68 minutes on the genetic map of K12. Since the gene encoding 2-keto-3-deoxyglucarate aldolase has not been mapped, it is not excluded that *f256* is this gene.

Catechol degradation



4-hydroxy-2-oxovalerate acetaldehyde pyruvate

Homoprotocatechuate degradation



2,4-dihydroxy-hept-2-ene-1,7-dioic acid succinate semialdehyde pyruvate

It is not known whether *orf243* is part of a *meta*-cleavage operon involved in aromatic catabolism by *E.coli* K12. The identity with highly conserved amino acid sequences of aldolases in *Pseudomonas* suggests that it may well have been a *meta*-cleavage enzyme at some stage during the evolution of K12, but whether the operon is still intact and active remains to be seen.

Chapter 5

Final Discussion

It seems clear that neither of the two open reading frames identified by hybridisation with oligonucleotide probes directed against a conserved sequence are likely to be new *E.coli* sigma genes. The minor codon probe, 746J, hybridised to the sense strand of *f229*, thus having sequence homology with the antisense strand. The major codon probe, 745J, did hybridise with *rpoD* and so seemed likely to be more successful in locating further sigma genes. However, the striking identity between the potential gene product of *orf243* and the family of *Pseudomonas* 4-hydroxy-2-oxovalerate aldolases suggests that this open reading frame encodes an isofunctional enzyme. Previous work in this laboratory has involved the use of anti-sigma peptide-antiserum to identify novel sigma proteins. Although this antiserum cross-reacted with σ^{70} and σ^{32}, characterisation of several other Sigma Cross-Reacting Proteins (SCRPs) indicated that they too were unlikely to be sigma factors (Ueshima *et al.*, 1992; Smillie, 1994).

The use of oligonucleotides to identify related sequences within a gene library has several disadvantages. Examination of hybrids formed between unique homologous probes and their cognate target sequences has revealed that short stretches of homology occuring by chance make a significant contribution to the hybridisation background (Lathe, 1985). The detremental effects of mismatches tend to be outweighed by the increased stability of hybrids formed by longer oligonucleotides, but a sufficiently long stretch of amino acid sequence is necessary. However, in this particular case the target sequence cannot be accurately defined, since the consensus sequence is a compromise between the sequences of several sigma factors. Furthermore, the target sequence has three leucines, two arginines, and a serine, all of which are encoded by six codons, thus increasing the uncertainty. Although σ^{70} does not have an identical sequence to the consensus core-binding sequence, 745J hybridised with *rpoD*. Less stringent washes might have resulted in hybridisation

with other known sigma genes, albeit at the expense of increased background. On the other hand, the more stringent washes might have revealed sigma genes having a subregion 2.1 sequence closely related to the consensus.

Although other known sigmas were not identified by either of the above approaches, this still raises doubts as to whether further members of the σ^{70} family are present in *E.coli*. Even if not, other unidentified sigma factors that are not related to the σ^{70} family may still exist. σ^{E} was first shown to function as a sigma factor by its ability to initiate transcription from the *rpoH* P3 promoter *in vitro* (Erickson and Gross, 1989). Although the amino acid sequence of σ^{E} does have homology with that of other sigmas of the σ^{70} family, it is less striking than the homology it shares with other members of the ECF subfamily (Lonetto *et al.,* 1994 and section 1.8.4). Other members of this subfamily differ so greatly in sequence from the canonical sigma factors that they are difficult to identify by standard sequence similarity searching methods. The sequence of *fecI* was published in 1990 (van Hove *et al.,* 1990) but FecI was not identified as a potential sigma factor. It was only implicated due to its membership of the ECF subfamily, then subsequently shown to function biochemically as a sigma factor (section 1.8.5).

Sigma factors have long been considered to control large regulons, but this may not necessarily be true. Some sigmas such as FecI may play an almost operon-specific role, analogous to that previously associated with certain transcriptional activators. This type of sigma may therefore have different properties and hence different protein domains from the more conventional sigmas, thus making it more difficult to identify them. These properties could include an ability to transduce a signal from a sensor before initiating transcription.

As more data become available, it may be possible to identify some novel sigma factors by sequence similarity alone. However, biochemical techniques may have to be employed to determine whether positive regulators of transcription can actually direct RNA synthesis in the absense of "known" sigma factors.

Bibliography

Abril, M.A., Michan, C., Timmis, K.N., and Ramos, J.L. (1989) *J. Bacteriol.* **171** : 6782-6790.

Altuvia, S., Almiron, M., Huisman, G., Kolter, R., and Storz, G. (1994) *Mol. Microbiol.* **13** : 265-272.

Anderson, P., and Roth, J. (1978) *J. Mol. Biol.* **119** : 147-166.

Angerer, A., Enz, S., Ochs, M., and Braun, V. (1995) *Mol. Microbiol.* **18** : 163-174.

Arnosti, D.N., and Chamberlin, M.J. (1989) *Proc. Natl. Acad. Sci. USA* **86** : 830-834.

Arnqvist, A., Olsen, A., and Normark, S. (1994) *Mol. Microbiol.* **13** : 1021-1032.

Asha, H., and Gowrishankar, J. (1993) *J. Bacteriol.* **175** : 4528-4537.

Atlung, T., and Brønsted, L. (1994) *J. Bacteriol.* **176** : 5414-5422.

Bachellier, S., Perrin, D., Hofnung, M., and Gilson, E. (1993) *Mol. Microbiol.* **7** : 537-544.

Bachellier, S., Saurin, W., Perrin, D., Hofnung, M., and Gilson, E. (1994) *Mol. Microbiol.* **12** : 61-70.

Bachman, B. (1972) *Bact. Rev.* **36** : 535-552.

Backman, K., Ptashne, M., and Gilbert, W. (1976) *Proc. Natl. Acad. Sci. USA* **73** : 4174-4178.

Balke, V.L. and Gralla, J.D. (1987) *J. Bacteriol.* **169** : 4499-4506.

Barany, F. (1985) *Gene* **37** : 111-123.

Barth, M., Marschall, C., Muffler, A., Fischer, D., and Hengge-Aronis, R. (1995) *J. Bacteriol.* **177** : 3455-3464.

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

Blaisdell, B.E., Rudd, K.E., Matin, A., and Karlin, A. (1993) *J. Mol. Biol.* **229** : 833-848.

Blattner, F.R., Burland, V., Plunkett III, G., Sofia, H.J., and Daniels, D.L. (1993) *Nucl. Acids Res.* **21** : 5408-5417.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994) *Cell* **78** : 889-896.

Blumenthal, H.J., and Fish, D.C. (1963) *Biochem. Biophys. Res. Comm.* **11** : 239-243.

van Bogelen, R.A., and Neidhardt, F.C. (1990) *Proc. Natl. Acad. Sci. USA* **87** : 5589-5593.

Boccard, F., and Prentki, P. (1993) *EMBO J.* **12** : 5019-5027.

Borukov, S., Lee, J., and Goldfarb, A. (1991) *J. Biol. Chem.* **266** : 23932-23935.

Botsford, J.L. (1990) *FEMS Microbiol. Lett.* **72** : 355-360.

Brandi, A., Pietroni, P., Gualerzi, C.O., and Pon, C.L. (1996) *Mol. Microbiol.* **19** : 231-240.

Braun, V. (1975) *Biochim. Biophys. Acta.* **415** : 335-377.

Brill, J.A., Quinlan-Walshe, C., and Gottesman, S. (1988) *J. Bacteriol.* **170** : 2599-2611.

Brønsted, L., and Atlung, T., (1994) *J. Bacteriol.* **176** : 5423-5428.

Broyles, S.S., and Pettijohn, D.E. (1986) *J. Mol. Biol.* **187** : 47-60.

Buck, M., and Cannon, W. (1992) *Nature* **358** : 422-424.

Buckle, M., Geiselman, J., Kolb, A., and Buc, H. (1991) *Nucl. Acids Res.* **19** : 883-840.

Burgess, R.R. (1969) *J. Biol. Chem.* **244** : 6168-6176.

Burgess, R.R., and Travers, A.A. (1970) *FASEB J.* **29** : 1164-1169.

Burlingame, R., and Chapman, P.J. (1983) *J. Bacteriol.* **155** : 113-121.

Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S., and Gross, C.A. (1981) *Nucl. Acids Res.* **9** : 2889-2903.

Cairney, J., Booth, I.R., and Higgins, C.F. (1985a) *J. Bacteriol.* **164** : 1218-1223.

Cashel, M., and Rudd, K.E. (1987) In : Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schlaechter, M., and Umbarger, H.E. (eds.) *E. coli* and *S. typhimurium* : Cellular and Molecular Biology. 1st. edit. pp1410-1438. American Society for Microbiology, Washington D.C.

Cashel, M., Gentry, D.R., Hernandez, V.J., and Vinella, D. (1995). In : Neidhardt, F.C. (editor-in-chief) *E. coli* and *S. typhimurium* : Cellular and Molecular Biology. 2nd. edit. in press. American Society for Microbiology, Washington D.C.

Chan, B., Minchin, S., and Busby, S. (1990) *FEBS Lett.* **267** : 46-50.

Chenchick, A., Babilashavi, R., and Mzabekov, A. (1981) *FEBS Lett.* **128** : 46-50.

Cooper, R.A., and Skinner, M.A. (1980) *J. Bacteriol.* **143** : 302-306.

Cooper, R.A., Jones, D.C.N., and Parrott, S. (1985) *J. Gen. Microbiol.* **131** : 2753-2757.

Craig, N.L., and Nash, H.A. (1984) *Cell* **39** : 707-716.

Craigie, R., Arndt-Jovin, D,J., and Mizuuchi, K. (1985) *Proc. Natl. Acad. Sci. USA* **82** : 7570-7574.

Csonka, L.N. (1989) *Microbiol. Rev.* **53** : 121-147.

Dagley, S. (1986) In : J.R. Sokatch (ed.) The Bacteria. vol. 10. The Biology of *Pseudomonas.* pp527-556. Academic Press, Inc. (London), Ltd., London.

Daniels, D., Zuber, P., and Losick, R. (1990) *Proc. Natl. Acad. Sci. USA* **87** : 8075-8079.

Deretic, V., Schurr, M.J., Boucher, J.C., and Martin, D.W. (1994) *J. Bacteriol.* **176** : 2773-2780.

Dersch, P., Kneip, S., and Bremer, E. (1994) *Mol. Gen. Genet.* **245** : 255-259.

Dimri, G.P., Rudd, K.E., Morgan, M.K., Bayat, H., and Ames, G.F-L. (1992) *J. Bacteriol.* **174** : 4583-4593.

Ding, Q., Kusano, S., Villarejo, M., and Ishihama, A. (1995) *Mol. Microbiol.* **16** : 649-656.

Dixon, N., and Kornberg, A. (1984) *Proc. Natl. Acad. Sci. USA* **81** : 424-428.

Doi, R.H., and Wang, L.F. (1986) *Microbiol. Rev.* **50** : 227-243.

Dombroski, A.J., Walter, W.A., Record, M.T., Siegele, D.A., and Gross, C.A. (1992) *Cell* **70** : 501-512.

Dombroski. A.J., Walter, W.A., and Gross, C.A. (1993) *Genes Dev.* **7** : 2446-2455.

Drlica, K., and Rouvière-Yaniv, J. (1987) *Microbiol. Rev.* **51** : 301-319.

Drlica, K. (1990) *Trends Genet.* **6** : 433-437.

Duetz, W.A., Marques, S., Dejong, C., Ramos, J.L., and Vanandel, J.G. (1994) *J. Bacteriol.* **176** : 2354-2361.

Enz, S., Braun, V., and Crosa, J.H. (1995) *Gene* **163** : 13-18.

Erickson, J.W., Vaughn, V., Walter, A.W., Neidhardt, F.C., and Gross, C.A. (1987) *Genes Dev.* **1** : 419-432.

Erickson, J.W., and Gross, C.A. (1989) *Genes Dev.* **3** : 1462-1471.

Espinosa-Urgel, M.E., and Tormo, A. (1993) *Nucl. Acids Res.* **21** : 3667-3670.

Forst, S., Delgado, J., and Inouye, M. (1989) *Proc. Natl. Acad. Sci. USA* **86** : 6052-6056.

Freundlich, M., Ramani, N., Mathew, E., Sirko, A., and Tsui, P. (1992) *Mol. Microbiol.* **6** : 2557-2563.

Fujita, N., Ishihama, A., Nagasawa, Y., and Ueda, S. (1987) *Mol. Gen. Genet.* **210** : 5-9.

Gaal, T., Ross, W., Blatter, E.E., Tang, H., Jia, X., Krishnan, V.V., Assa-Munt, N., Ebright, R.H., and Gourse, R.L. (1996) *Genes Dev.* **10** : 16-26.

Gallegos, M.T., Michan, C., and Ramos, J.L. (1993) *Nucl. Acids Res.* **21** : 807-810.

Gamer, J., Bujard, H., and Bukau, B. (1992) *Cell* **69** : 833-842.

Gamer, J., Multhaup, G., Tomoyasu, Y., McCarty, J.S., Rudiger, S., Schonfeld, H.J., Schirra, C., Bujard, H., and Bukau, B. (1996) *EMBO J.* **15** : 607-617.

Gardella, J., Moyle, H., and Susskind, M.M. (1989) *J. Mol. Biol.* **206** : 579-596.

Gentry, D.R., Hernandez, V.J., Nguyen, L.H., Jensen, D.B., and Cashel, M. (1993) *J. Bacteriol.* **175** : 7982-7989.

Gibson, T.J. (1984) Ph.D Thesis, University of Cambridge, U.K.

Gillen, K.L., and Hughes, K.T. (1991) *J. Bacteriol.* **173** : 2301-2310.

Gilson, E., Clément, J.M., Brutlag, D., and Hofnung, M.. (1984) *EMBO J.* **3** : 1417-1421.
Gilson, E., Rousset, J.P., Clément, J.M., and Hofnung, M. (1986) *Ann. Microbiol. (Inst. Pasteur)* **137B** : 259-270.

Gilson, E., Clément, J.M., Perrin, D., and Hofnung, M. (1987) *Trends Genet.* **3** : 226-230.

Gilson, E., Bachellier, S., Perrin, S., Perrin, D., Grimont, P.A.D., Grimont, F., and Hofnung, M. (1990) *Res. Microbiol.* **141** : 1103-1116.

Gilson, E., Saurin, W., Perrin, D., Bachellier, S., and Hofnung, M. (1991) *Nucl. Acids Res.* **19** : 1375-1383.

Glass, R.E., and Nene, V. (1985) In : Glass, R.E., and Spizek, J. (eds.). Gene Manipulation and Expression p155. Croom Helm, London..

Glass, R.E., Jones, S.T., and Ishihama, A. (1986a) *Mol. Gen. Genet.* **203** : 265-268.

Glass, R.E., Honda, A., and Ishihama, A. (1986b) *Mol. Gen. Genet.* **203** : 492-495.

Glass, R.E., Jones, S.T., Nomura, T., and Ishihama, A. (1987) *Mol. Gen. Genet.* **210** : 1-4.

Goldenberg, D., Azar, I., and Oppenheim, A.B. (1996) *Mol. Microbiol.* **19** : 241-248.

Goldstein, J., Pollitt, N.S., and Inouye, M. (1990) *Proc. Natl. Acad. Sci. USA* **87** : 283-287.

Gopal, V., Ma, H.W., Kumaran, M.K., and Chatterji, D. (1994) *J. Mol. Biol.* **242** : 9-22.

Gowrishankar, J. (1985) *J. Bacteriol.* **164** 434-445.

Gribskov, M., and Burgess, R.R. (1986) *Nucl. Acids Res.* **14** : 6745-6763.

Gross, C.A., Straus, D.B., Erickson, J.W., and Yura, T. (1990) In : Morimoto, R.I., Tissieres, A., and Georgopolous, C. (ed.) Stress Proteins in Medicine pp162-187. Cold Spring Harbour Laboratory press, Cold Spring Harbour, N.Y.

Gross, C.A., Lonetto, M., and Losick, R. (1992) In : Yamamoto, K., and McNight, S. (ed.) Control of Transcription. Sigma Factors. Cold Spring Harbour Laboratory press, Cold Spring Harbour, N.Y.

Gualerzi, C.O., Losso, M.A., Lammi, M., Friedrich, K., Pawlik, R.T., Canonaco, M.A., Gianfranceschi, A., Pingoud, A., and Pon, C.L. (1986) In : Gualerzi, C.O., and Pon, C.L. (ed.) Bacterial Chromatin pp101-134. Springer-Verlag, Heidelberg, FRD.

Guoy, M., and Gautier, C. (1982) *Nucl. Acids Res.* **10** : 7055-7074.

Gussin, G.N., Ronson, C.W., and Ausubel, F.M. (1986) *Annu. Rev. Genet.* **20** : 567-591.

Harayama, S., and Rekik, M. (1993) *Mol. Gen. Genet.* **239** : 81-89.

deHaseth, P.L., and Helmann, J.D. (1995) *Mol. Microbiol.* **16** : 817-824.

Hawley, D., and McClure, W.R. (1983) *Nucl. Acids Res.* **11** : 2237-2255.

Hayward, R.S., Igarashi, K., and Ishihama, A. (1991) *J. Mol. Biol.* **221** : 23-29.

Hayward, R.S., Kumar, A., Grimes, B., and Logan, M. (1992) *J. Cell Biochem.* (suppl.) **16E** : 141.

Heisler, L.M., Suzuki, H, Landick, R, and Gross, C.A. (1993) *J. Biol. Chem.* **268** : 25369-25375.

Heisler, L.M., Feng, G.H., Jin, D.J., Gross, C.A., and Landick, R. (1996) *J. Biol. Chem.* **271** : 14572-14583.

Helmann, J.D., and Chamberlin, M.J. (1988) *Annu. Rev. Biochem.* **57** : 839-872.

Helmann, J.D. (1991) *Mol. Microbiol.* **5** : 2875-2882.

Hendrick, J.P., and Hartl, F.U. (1993) *Annu. Rev. Biochem.* **62** : 349-384.

Hengge, R. and Boos, W. (1983) *Biochim. Biophys. Acta.* **737** : 443-478.

Hengge-Aronis, R. (1993) *Cell* **72** : 165-168.

Hengge-Aronis, R. Lange, R., Henneborg, N., and Fischer, D. (1993) *J. Bacteriol.* **173** : 7918-7924.

Herman, B.G., and Frischauf, A.M. (1987) *Methods in Enzymology* **52** : 180-182.

Herman, N.D., and Schneider, T.D. (1992) *J. Bacteriol.* **174** :3558-3560.

Hernandez, V.J., and Cashel, M. (1995) *J. Mol. Biol.* **252** : 536-549.

Hess, F.F., Bourret, R.B., and Simon, M.I. (1988) *Nature* **336** : 139-143.

Higgins, C.F., Ames, G.F-L., Barnes, W.M., Clément, J.M., and Hofnung, M. (1982) *Nature* **298** : 760-762.

Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G., and Bremer, E. (1988a) *Cell* **52** : 569-584.

Higgins, C.F., McLaren, R.S., and Newbury, S.F. (1988b) *Gene* **72** : 3-14.

Hofer, B., Backhaus, S., and Timmis, K.N. (1994) *Gene* **144** : 9-16.

van Hove, B., Staudenmaier, H., and Braun, V. (1990) *J. Bacteriol.* **172** : 6749-6758.

Hsieh, L.S., Rouvière-Yaniv, J., and Drlica, K. (1991) *J. Bacteriol.* **173** : 3914-3917.

Hsieh, L.S., and Gralla, J.D. (1994) *J. Mol. Biol.* **239** : 15-24.

Hsieh, L.S., Tintut, Y., and Gralla, J.D. (1994) *J. Biol. Chem.* **269** : 373-378.

Igarashi, K., Fujita, N., and Ishihama, A. (1991) *J. Mol. Biol.* **218** : 1-6.

Ikemura, T. (1981) *J. Mol. Biol.* **151** : 389-409.

Ikemura, T. (1985) *J. Mol. Evol.* **2** : 13-24.

Inouye,S., Wang, S., Sekizawa, J., Halegoua, S., and Inouye, M. (1977) *Proc. Natl. Acad. Sci. USA* **74** : 1004-1008.

Inouye, S., Nakazawa, A., and Nakazawa, T. (1988) *Gene* **66** : 301-306.

Ishihama, A. (1981) *Advan. Biophys.* **14** : 1-35.

Ishihama, A. (1993) *J. Bacteriol.* **175** : 2483-2489.

Iwakuri, Y., Ishihama, A., and Yura, T. (1973) *Mol. Gen. Genet.* **121** : 181-196.

Jaworski, A., Higgins, N.P., Wells, R.D., and Zacharias, W. (1991) *J. Biol. Chem.* **266** : 2576-2581.

Jenkins, J.R. (1987) Ph.D. Thesis, University of Leicester, U.K.

Jenkins, J.R., and Cooper, R.A. (1988) *J. Bacteriol.* **170** : 5317-5324.

Jenkins, D.E., Chaison, S.A., and Matin, A. (1990) *J. Bacteriol.* **172** : 2779-2781.

Jin, D.J., and Gross, C.A. (1988) *J. Mol. Biol.* **202** : 45-58.

Johnson, R.C., Bruist, M.F., and Simon, M.I. (1986) *Cell* **46** : 531-539.

Jones, P.G., van Bogelen, R.A., and Neidhardt, F.C. (1987) *J. Bacteriol.* **169** : 2092-2095.

Jones, P.G., Cashel, M., Glaser, G., and Neidhardt, F.C. (1992a) *J. Bacteriol.* **174** : 3903-3914.

Jones, P.G., Krah, R., Tafuri, S.R., and Wolffe, A.P. (1992b) *J. Bacteriol.* **174** : 5798-5802.

Jones, P.G. and Inouye, M. (1994) *Mol. Microbiol.* **11** : 811-818.

Jones, P.G., Mitta, M., Kim, Y., Jiang, W.N., and Inouye, M. (1996) *Proc. Natl. Acad. Sci. USA* **93** : 76-80.

Juang, Y.L., and Helmann, J.D. (1994) *J. Mol. Biol.* **235** : 1470-1488.

Jung, J.U., Gutierrez, C., and Villarejo, M.R. (1989) *J. Bacteriol.* **171** : 511-520.

Jung, J.U., Gutierrez, C., Martin, F., Ardourel, M., and Villarejo, M. (1990) *J. Biol. Chem.* **265** : 10574-10581.

Keilty, S., and Rosenberg, M. (1987) *J. Biol. Chem.* **262** : 6389-6395.

Kessler, B., Marques, S., Kohler, T., Ramos, J.L., Timmis, K.N., and Delorenzo, V. (1994) *J. Bacteriol.* **176** : 5578-5582.

Kikuchi, Y., Yasukochi, Y., Nagata, Y., Fukuda, M., and Takagi, M. (1994) *J. Bacteriol.* **176** : 4269-4276.

Kim, S.K., Makino, K., Amemura, M., Nakata, A., and Shinagawa, H. (1995) *Mol. Gen. Genet.* **248** : 1-8.

Kimura, M., Fujita, N., and Ishihama, A. (1994) *J. Mol. Biol.* **242** : 107-115.

Kimura, M., and Ishihama, A. (1995a) *J. Mol. Biol.* **248** : 756-767.

Kimura, M., and Ishihama, A. (1995b) *J. Mol. Biol.* **254** : 342-349.

Kohara, Y., Akiyama, K., and Isono, K. (1987) *Cell* **50** : 495-508.

Kolter, R., Siegele, D.A., and Tormo, A. (1993) *Annu. Rev. Microbiol.* **47** : 855-874.

Kornberg, A. (1980) *DNA Replication* 101-110.

Kulakaukas, S., Wikstrom, P.M., and Berg, D.E. (1991) *J. Bacteriol.* **173** : 2633-2638.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A., and Hayward, R.S. (1993) *J. Mol. Biol.* **232** : 406-418.

Kumar, A., Grimes, B., Fujita, N., Makino, K., Malloch, R.A., Hayward, R.S, and Ishihama, A. (1994) *J. Mol. Biol.* **235** : 405-413.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A., and Hayward, R.S. (1993) *J. Mol. Biol.* **232** : 406-418.

Kumar, A., Grimes, B., Fujita, N., Makino, K., Malloch, R.A., Hayward, R.S, and Ishihama, A. (1994) *J. Mol. Biol.* **235** : 405-413.

Kumar, A., Grimes, B., Logan, M., Wedgwood, S., Williamson, H., and Hayward, R.S. (1995a) *J. Mol. Biol.* **246** : 563-571.

Kumar, A., Williamson, H.S., Fujita, N., Ishihama, A., and Hayward, R.S. (1995b) *J. Bacteriol.* **177** : 5193-5196.

Kur, J., Hasan, N., and Szybalski, W. (1989) *Gene* **81** : 1-15.

Kusano, S., Ding, Q., Fujita, N., and Ishihama, A. (1996) *J. Biol. Chem.* **271** : 1998-2004.

Kustu, S., Santero, E., Keener, J., Popham, D., and Weiss, D. (1989) *Microbiol. Rev.* **53** : 367-376.

Laemmli, U.K. (1970) *Nature* **227** : 680-685.

Laimins, L.A., Rhoads, D.B., and Epstein, W. (1981) *Proc. Natl. Acad. Sci. USA* **78** : 464-468.

Lange, R., and Hengge-Aronis, R. (1994) *Genes Dev.* **8** : 1600-1612.

La Teana, A., Brandi, A., Falconi, M., Spurio, R., and Pon, C.L. (1991) *Proc. Natl. Acad. Sci. USA* **88** : 10907-10911.

Lathe, R. (1985) *J. Mol. Biol.* **183** : 1-12.

Lau, P.C.K., Bergenon, H., Labbe, D., Wang, Y., Brousseau, R., and Gibson, D.T. (1994) *Gene* **146** : 7-13.

Lee, H.S., Ishihama, A., and Kustu, S. (1993) *J. Bacteriol.* **175** : 2479-2482.

Lemmi. M., Paci, M., Pon, C.L., Losso, M.A., Miano, A., Pawlik, R.T., Gianfranceschi, G.L., and Gualerzi, C.O. (1984) In : Hubscher, H., and Spadori, S. (ed.) Proteins Involved in DNA Replication pp467-478. Plenum Press, N.Y.

Lesley, S.A., and Burgess, R.R. (1989) *Biochemistry* **28** : 7728-7734.

Liberek, K., Galitski, T.P., Zylicz, M., and Georgopoulos, C. (1992) *Proc. Natl. Acad. Sci. USA* **89** : 5316-5320.

Lipinska, B., Sharma, S., and Georgopoulos, C. (1988) *Nucl. Acids Res.* **16** : 10053-10067.

Liu, K., Zhang, Y., Severinov, K., Das, A., and Hanna, M.M. (1996) *EMBO J* **15** : 150-160.

Liut, X.Y., Fujita, N., Ishihama, A., and Matsumura, P. (1995) *J. Bacteriol.* **177**: 5186-5188.

Lomovskaya, O.L., Kidwell, J.P., and Matin, A. (1994) *J. Bacteriol.* **176** : 3928-3935.

Lonetto, M.A., Gribskov, M., and Gross, C.A. (1992) *J. Bacteriol.* **174** : 3843-3849.

Lonetto, M.A., Brown, K.L., Rudd, K.E., and Buttner, M.J. (1994) *Proc. Natl. Acad. Sci. USA* **91** : 7573-7577.

Lucht, J.M., Dersch, P., Kempf, B., and Bremer, E. (1994) *J. Biol. Chem.* **269** : 6578-6586.

Lund, E., and Kjeldgaard, N.O. (1972) *Eur. J. Biochem.* **28** : 316-326.

Mackow, E.R., and Chang, F.N. (1983) *Mol. Gen. Genet.* **192** : 5-9.

Magasanik, B. (1982) *Annu. Rev. Genet.* **16** : 135-168.

Magasanik, B. (1988) *Trends Biol. Sci.* **13** : 475-479.

Makino, K., Amemura, M., Kim, S.K., Nakata, A., and Shinagawa, H. (1993) *Genes Dev.* **7** : 149-160.

Makino, K., Amemura, M., Kawamoto, T., Kimura, S., Shinagawa, H., Nakata, A., and Suzuki, M. (1996) *J. Mol. Biol.* **259** : 15-26.

Manna, D., and Gowrishankar, J. (1994) *J. Bacteriol.* **176** : 5378-5384.

Manoil, C., and Beckwith, J. (1985) *Proc. Natl. Acad. Sci. USA* **83** : 8129-8133.

Manoil, C., and Beckwith, J. (1986) *Science* **233** : 1403-1408.

Manson, M.D., Boos, W., Bassford, P.J., and Rasmussen, B.A. (1985) *J. Biol. Chem.* **260** : 9727-9733.

Margolis, P.M., Driks, A., and Losick, R. (1991) *Science* **248** : 562-565.

Marques, S., and Ramos, J.L. (1993) *Mol. Microbiol.* **9** : 923-929.

May, G., Faatz, E., Villarejo, M., and Bremer, E., (1986) *Mol. Gen. Genet.* **205** : 225-233.

May, G., Faatz, E., Lucht, J.M., Haardt, M., Bolliger, M. and Bremer, E. (1989) *Mol. Microbiol.* **3** : 1521-1531.

McCann, M.P., Kidwell, J.P., and Matin, A. (1991) *J. Bacteriol.* **173** : 4188-4194.

McCann, M.P., Fraley, C.D., and Matin, A. (1993) *J. Bacteriol.* **175** : 2143-2149.

McClure, W.R. (1985) *Annu. Rev. Biochem.* **54** : 171-204.

McCarty, J.S., Rudiger, S, Schonfeld, H.J., Schneider-Mergener, J., Nakahigashi, K., Yura, T., and Bukau, B. (1996) *J. Mol. Biol.* **256** : 829-837.

Mecsas, J., Rouvière, P.E., Erickson, J.W., Donohue, T.J., and Gross, C.A. (1993) *Genes Dev.* **7** : 2618-2628.

Meinkoth, J., and Wahl, G. (1984) *Anal. Biochem.* **138** : 267-284.

Mellies, J., Brems, R., and Villarejo, M. (1994) *J. Bacteriol.* **176** : 3638-3645.

Merrick, M., Gibbins, J., and Toukdarian, A. (1987) *Mol. Gen. Genet.* **210** : 323-330.

Merrick, M.J. (1993) *Mol. Microbiol.* **10** : 903-909.

Michaelis,S., and Beckwith, J. (1982) *Annu. Rev. Microbiol.* **36** : 435-465.

Mukhopadhyay, S., and Schellhorn, H.E. (1994) *J. Bacteriol.* **176** : 2300-2307.

Mulvey, M., and Loewen, P.C. (1989) *Nucl. Acids Res.* **17** : 9979-9991.

Mustaev, A., Kashlev, M., Lee, J.Y., Polyakov, A., Lebedev, A., Zalenskaya, K., Grachev, M., Goldfarb, A., and Nikiforov, V. (1991) *J. Biol. Chem.* **266** : 23927-23931.

Nagai, H., Yuzawa, H., and Yura, T. (1991) *Proc. Natl. Acad. Sci. USA* **88** : 10515-10519.

Nagai, H., Yuzawa, H., Kanemori, M., and Yura, T. (1994) *Proc. Natl. Acad. Sci. USA* **91** : 10280-10284.

Naito, A., Naito, S., and Ikeda, H. (1984) *Mol. Gen. Genet.* **193** : 238-243.

Neidhardt, F.C., van Bogelen, R.A., and Vaughn, V. (1984) *Annu. Rev. Genet.* **18** : 295-329.

Nene, V., and Glass, R.E. (1984) *Mol. Gen. Genet.* **194** :166-172.

Newbury, S., Smith, N.H., and Higgins, C.F. (1987) *Cell* **51** : 1131-1143.

Ninfa, A.J., and Bennett, R.L. (1991) *J. Biol. Chem.* **266** : 6888-6893.

Noda, A., Courtright, J.B., Denor, P.F., Webb, G., Kohara, Y., and Ishihama, A. (1991) *Biotechniques* **10** : 474.

Ochs, M., Veitinger, S., Kim, I., Welz, D., Angerer, A., and Braun, V. (1995) *Mol. Microbiol.* **15** : 119-132.

Ochs, M., Angerer, A., Enz, S., and Braun, V. (1996) *Mol. Gen. Genet.* **250** : 455-465.

Oen, H., and Wu, C.W. (1978) *Proc. Natl. Acad. Sci. USA* **75** : 1778-1782.

Ogawa, T., Baker, T.A., van der Ende, A., and Kornberg, A. (1985) *Proc. Natl. Acad. Sci. USA* **82** : 3562-3566.

Ohnishi, K., Kutsukake, K., Suzuki, H., and Iino, T. (1990) *Mol. Gen. Genet.* **221** : 139-147.

Ohno, S. (1970) In : Evolution by Gene Duplication. Springer-Verlag, N.Y.

Ohshima, A., Inouye, S., and Inouye, M. (1992) *Proc. Natl. Acad. Sci. USA* **89** : 1016-1020.

Ollis, P.C., and White, S.W. (1987) *Chem. Rev.* **87** : 981-995.

Ophir, T., and Gutnick, D.L. (1994) *Appl. Environ. Microbiol.* **60** : 740-745.

Oppenheim, A.B., Rudd, K.E., Mendelson, I, and Teff, D. (1993) *Mol. Microbiol.* **10** : 113-122.

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

Ovchinnikov, Y.A., Monastryskaya, G.S., Gubanov, V.V., Guryev, S.O., Salomatina, I.S., Shuvaeva, T.M., Lipkin, V.M., and Svedlov, E.D. (1982) *Nucl. Acids Res.* **10** : 4035-4044.

Owen-Hughes, T.A., Pavitt, G.D., Santos, D.S., Sidebotham, J.M., Hulton, C.S, Hinton, J.C.D., and Higgins, C.F. (1992) *Cell* **71** : 255-265.

Ozaki, M., Wada, A., Fujita, N., and Ishihama, A. (1991) *Mol. Gen. Genet.* **230** : 17-23.

Ozaki, M., Fujita, N., Wada, A., and Ishihama, A. (1992) *Nucl. Acids Res.* **20** : 257-261.

Painbeni, E., Mouray, E., Gottesman, S., and Rouvière-Yaniv, J. (1993) *J. Mol. Biol.* **234** : 1021-1037.

Pao, C.C., and Dyess, B.T. (1981) *J. Biol. Chem.* **256** : 2252-2257.

Parker, C.T., Kloser, A.W., Schnaitman, C.A., Stein, M.A., Gottesman, S., and Gibson, B.W. (1992) *J. Bacteriol.* **174** : 2525-2538.

Parrott, S., Jones, S., and Cooper, R.A. (1987) *J. Gen. Microbiol.* **133** : 347-353.

Pettijohn, D.E. (1982) *Cell* **30** : 667-669.

Poniggia, A., Negri,. A., Beltrame, M., and Bianchi, M.E. (1993) *Mol. Microbiol.* **7** : 343-350.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989) *Science* **243** : 629-635.

Post, L.E., and Nomura, M. (1979) *J. Biol. Chem.* **254** : 10604-10606.

Qoronfleh, M.W., Debouck, C., and Keller, J. (1992) *J. Bacteriol.* **174** : 7902-7909.

Raina, S., Missiakas, D., and Georgopoulos, C. (1995) *EMBO J.* **14** : 1043-1055.

Ramos, J.L., Rojo, F., Zhou, L., and Timmis, K.N. (1990) *Nucl.* Acids *Res.* **18** : 2149-2152.

Reitzer, L.J., and Magasanik, B. (1986) *Cell* **45** : 785-792.

Richetti, M., and Buc, H. (1993) *EMBO J.* **12** : 387-396.

Richey, B., Cayley, D.S., Mossing, M.C., Kolka, C., Anderson, C.F., Farrar, T.C., and Record, M.T. (1987) *J. Biol. Chem.* **262** : 7157-7164.

Roland, G.C., and Glass, R.E. (1990) *BioEssays* **12** : 343-346.

Ronson, C.W., Nixon, B.T., and Ausubel, F.M. (1987) *Cell* **49** : 579-581.

Roper, D.I., Fawcett, T., and Cooper, R.A. (1993) *Mol. Gen. Genet.* **237** 241-250.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., and Ishihama, A. (1993) *Science* **262** : 1407-1413.

Rouvière, P.E., De Las Peñas, A., Mecsas, J., Lu, C.Z., Rudd, K.E., and Gross, C.A. (1995) *EMBO J.* **14** : 1032-1042.

Sagitov, V., Nikiforov, V. and Goldfarb, A. (1993) *J. Biol. Chem.* **268** : 2195-2202.

Salser, W., Gesteland, R.F., and Bolle, A. (1967) *Nature* **215** : 588-591.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual. 2nd edition. Cold Spring Harbour Laboratory press.

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

Sanzey, B. (1979) *J. Bacteriol.* **138** : 40-47.

Sasse-Dwight, S., and Gralla, J.D. (1988) *Proc. Natl. Acad. Sci. USA* **85** : 8934-8938.

Schweder, T., Lee, K.H., Lomovskaya, O., and Matin, A. (1996) *J. Bacteriol.* **178** : 470-476.

Schultz-Hauser, G., van Hove, B., and Braun, V. (1992) *FEMS Microbiol. Lett.* **95** 231-234.

Severinov, K., Soushko, M., Goldfarb, A. and Nikiforov, V. (1993) *J. Biol. Chem.* **268** : 14820-14825.

Severinov, K., Soushko, M., Goldfarb, A. and Nikiforov, V. (1994) *Mol. Gen. Genet.* **244** : 120-126.

Severinov, K., Markov, D., Severinova, E., Nikiforov, V., Landick, R., Darst, S.A., and Goldfarb, A. (1995) *J. Biol. Chem.* **270** : 23926-23929.

Shaw, D.J., and Guest, J.R. (1982) *J. Gen. Microbiol.* **128** : 2221-2228.

Shimitake, H., and Rosenberg, M. (1981) *Nature* **292** : 128-132.

Shingler, V., Powlowski, J., and Marklund, Y. (1992) *J. Bacteriol.* **174** : 711-724.

Shingler, V., Bartilson, M., and Moore, T. (1993) *J. Bacteriol.* **175** : 1596-1604.

Shorenstein, R.G., and Losick, R. (1973) *J. Biol. Chem.* **248** : 6170-6173.

Shyamala, V., Schneider, E., and Ames, G.F-L. (1990) *EMBO J.* **9** 939-946.

Siegele, D.A., Hu, J.C., Walter, W.A., and Gross, C.A. (1989) *J. Mol. Biol.* **206** : 591-603.

Siegele, D.A., and Kolter, R. (1992) *J. Bacteriol.* **174** : 345-348.

Simpson, R.B. (1979) *Cell* **18** : 277-285.

Sledjeski, D.D., and Gottesman, S. (1995) *Proc. Natl. Acad. Sci. USA* **92** : 2003-2007.

Sledjeski, D.D., and Gottesman, S. (1996) *J. Bacteriol.* **178** : 1204-1206.

Smillie, D.A., Hayward, R.S., Suzuki, T., Fujita, N., and Ishihama, A. (1992) *J. Bacteriol.* **174** : 3826-3827.

Smillie, D.A. (1994) Ph.D Thesis, University of Edinburgh, U.K.

Spurio, R., Dürrenberger, M., Falconi, M., La Teana, A., Pon, C.L., and Gualerzi, C.O. (1992) *Mol. Gen. Genet.* **231** : 201-211.

Stern, M.J., Ames, G. F-L., Smith, N.H., Robinson, E.C., and Higgins, C.J. (1984) *Cell* **37** : 1015-1026.

Stern, M.J., Prossnitz, E., and Ames, G.F-L. (1988) *Mol. Microbiol.* **2** : 141-152.

Stirling, D.A., Hulton, C.S.J., Waddell, L., Park, S.F., Stewart, G.S.A.B., Booth, I.R., and Higgins, C.F. (1989) *Mol. Microbiol.* **3** : 1025-1038.

Stout, V., and Gottesman, S. (1990) *J. Bacteriol.* **172** : 659-669.

Stout, V., Torres-Cabassa, A., Maurizi, M.R., Gutnick, D., and Gottesman, S. (1991) *J. Bacteriol.* **173** : 1738-1747.

Straus, D.B., Walter, W.A., and Gross, C.A. (1987) *Nature* **329** : 348-351.

Straus, D.B., Walter, W.A., and Gross, C.A. (1990) *Genes Dev.* **4** : 2202-2209.

Stringfellow, J.M., Turpin, B., and Cooper, R.A. (1995) *Gene* **166** : 73-76.

Studier, F.W., and Moffatt, B.A. (1986) *J. Mol. Biol.* **189** : 113-130.

Studier, F.W. (1990) *Methods in Enzymology* **185** : 60.

Sutherland, L., Cairney, J., Elmore, M.J., Booth, I.R., and Higgins, C.F. (1986) *J. Bacteriol.* **168** : 805-814.

Tanabe, H., Goldstein, J., Yang, M., and Inouye, M. (1992) *J. Bacteriol.* **174** : 3867-3875.

Tanaka, K., Shiina, T., and Takahashi, H. (1988) *Science* **242** : 1040-1042.

Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A., and Takahashi, H. (1993) *Proc. Natl. Acad. Sci. USA* **190** : 3511-3515.

Tanaka, K., Kusano. S., Fujita, N., Ishihama, A., and Takahashi, H. (1995) *Nucl. Acids Res.* **23** : 827-834.

Tavorima, P.L., Reznikoff, W.S., and Gross, C.A. (1996) *J. Mol. Biol.* **258** : 213-223.

Tilly, K., McKittrick, N., Zylicz, M., and Georgopoulos, C. (1983) *Cell* **34** : 641-646.

Tintut, Y., and Gralla, J.D. (1995) *J. Bacteriol.* **177** : 5818-5825.

Tintut, Y., Wang, J.T., and Gralla, J.D. (1995) *Genes Dev.* **9** : 2305-2313.

Tolner, B., Poolman, B., Wallace, B., and Konings, W.N. (1992) *J. Bacteriol.* **174** : 2391-2393.

Tomoyasu, T., Gamer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, A.J., Oppenheim, A.B., Yura, T., Yamanaka, K., Niki, H., Hiraga, S., and Ogura, T. (1995) *EMBO J.* **14** : 2551-2560.

Torres-Cabassa, A.S., and Gottesman, S. (1987) *J. Bacteriol.* **169** : 981-989.

Tupper, A.E., Owen-Hughes, T.A., Ussery, D.W., Santos, D.S., Ferguson, D.J.P., Sidebotham, J.M., Hinton, J.C.D., and Higgins, C.F. (1994) *EMBO J.* **13** : 258-268.

Ueguchi, C., and Mizuno, T. (1993) *EMBO J.* **12** : 1039-1046.

Ueshima, R., Fujita, N., and Ishihama, A. (1992) *Biochem. Biophys. Res. Comm.* **184** : 634-639.

Wada, A., Igarashi, K., Yoshimura, S., Aimoto, S., and Ishihama, A. (1995) *Biochem. Biophys. Res. Comm.* **214** : 410-417.

Wang, Q.P., and Kaguni, J.M. (1989) *J. Bacteriol.* **171** : 4248-4253.

Waldburger, C., Gardella, T., Wong, R., and Susskind, M.M. (1990) *J. Mol. Biol.* **215** : 267-276.

Waldburger, C., and Susskind, MM. (1994) *J. Mol. Biol.* **235** : 1489-1500.

Walderhaug, M.O., Polarek, J.W., Voelkner, P., Daniel, J.M., Hesse, J.E., Altendorf, K., and Epstein, W. (1992) *J. Bacteriol.* **174** 2152-2159.

Wallace, B., Yang, Y., Hong, J., and Lum, D. (1990) *J. Bacteriol.* **172** : 3214-3220

Weilbacher, R., Hebron, C., Feng, G., and Landick, R. (1994) *Genes Dev.* **8** : 2913-2927.

Wistow, G. (1990) *Nature* **344** : 823-824.

Wolffe, A.., Tafuri, S., Ranjan, M., and Familiari, M. (1992) *N. Biologist* **4** : 290-298.

Wolffe, A. (1994) *BioEssays* **16** : 245-250.

Wong, C., Tintut, Y., and Gralla, J.D. (1994) *J. Mol. Biol.* **236** : 81-90.

Worsey, M.J., and Williams, P.A. (1975) *J. Bacteriol.* **124** : 7-13.

Wylie, D., Stock, A., Wong, C.Y., and Stock, J. (1988) *Biochem. Biophys. Res. Comm.* **151** : 891-896.

Xiao,H., Kalman, M., Ikehara, K., Zemel, S., Glaser, G., and Cashel, M. (1991) *J. Biol. Chem.* **266** : 5980-5990.

Yamanaka, K., Mitani, T., Ogura, T., Niki, H., and Hiraga, S. (1994) *Mol. Microbiol.* **13** : 301-312.

Yamashino, T., Ueguchi, C., and Mizuno, T. (1995) *EMBO J.* **14** : 594-602.

Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D., and Somero, G.N. (1982) *Science* **217** : 1214-1222.

Yang, Y., and Ames, G.F-L. (1988) *Proc. Natl. Acad. Sci. USA* **85** : 8850-8854.

Yim, H.H., and Villarejo, M. (1992) *J. Bacteriol.* **174** : 3637-3644.

Yim, H.H., Brems, R.L., and Villarejo, M. (1994) *J. Bacteriol.* **176** : 100-107.

Yoshida, T., Ueguchi, C., Yamada, H., and Mizuno, T. (1993) *Mol. Gen. Genet.* **237** : 113-122.

Youderian, P., Bouvier, S., and Susskind, M.M. (1982) *Cell* **30** : 843-853.

Yura, T., and Ishihama, A. (1979) *Annu. Rev. Genet.* **13** : 59-97.

Yura, T., Nagai, H., and Mori, H. (1993) *Annu. Rev. Microbiol.* **47** : 321-350.

Yuzawa, H., Nagai, H., Mori, H., and Yura, T. (1993) *Nucl. Acids Res.* **21** : 5449-5455.

Zhou, Y.N., Walter, W.A., and Gross, C.A. (1992) *J. Bacteriol.* **174** : 5005-5012.

Zimmermann, L., Hantke, K., and Braun, V. (1984) *J. Bacteriol.* **159** : 271-277.

Zinoni, F., Birkmann, A., Stadtman, T.C., and Bock, A. (1986) *Proc. Natl. Acad. Sci. USA* **83** : 4650-4654.

Zuber, F., Kotlarz, D., Rimsky, S., and Buc, H. (1994) *Mol. Microbiol.* **12** : 231-240.

Zuber, P., Healy, J., Carter III, H.L., Cutting, S., Moran, C.P., and Losick, R. (1989) *J. Mol. Biol.* **206** : 605-614.