I

Genes Encoding Sigma Cross-Reacting Proteins of *Escherichia coli*

DAVID ANDREW SMILLIE

A thesis presented for the degree of Ph.D.

Institute of Cell and Molecular Biology

University of Edinburgh

June 1994

DECLARATION

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

ACKNOWLEDGEMENTS

I wish to thank a number of people who have been of help to me during my period of study: I am deeply grateful to Richard Hayward for the thoughtful guidance and encouragement he has offered and the many social gatherings he has instigated during my time in his laboratory. Many thanks to the rest of the crew (past and present); Ashok, Richie, Brenda, Tom, Steve and Helen who have contributed towards a stimulating and pleasant working environment. Thanks also to Sarah McQuay for helping with sequence alignments and to Alex Paton for advice on other computing matters. I owe a great deal to my family for all their support (moral and financial); a special word for my mother who has spent a great deal of her spare time typing this thesis. Most of all, I thank Eirlys for everything.

ABBREVIATIONS

aa	amino acid(s)
Amp $^{r/s}$	ampicillin resistant (r) or sensitive (s)
AMPS	ammonium persulphate
bp	base pair(s)
BSA	bovine serum albumin
CHP	cumene hydroperoxide
cDNA	complementary DNA
$Cm^{r/s}$	chloramphenicol resistant (r) or sensitive (s)
ddNTP	dideoxynucleoside 5'-triphosphate
dNTP	deoxynucleoside 5'-triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
dH_2O	distilled water
EDTA	diamino ethane tetraacetic acid
EGTA	ethylene glycol-bis (beta-amino-ethyl ether) N, N, N', N'-tetra acetic acid
EtdBR	ethidium bromide
g	acceleration due to gravity
Kan^r	kanamycin resistant
kb	kilobase
IPTG	isopropyl-β-D-thiogalactoside
LB	Luria broth
MD	menadione
MOI	multiplicity of infection
MOPS	3-[N-morpholino] propane sulphonic acid

mRNA	messenger RNA
nt	nucleotide(s)
NUV	near ultra violet
PL	plumbagin
PEG	polyethylene glycol
p.f.u.	plaque forming unit(s)
PQ	paraquat
psi	pounds per square inch
RF	replicative form
RNA	ribonucleic acid
RNase	ribonuclease
ss	single-stranded
TEMED	N,N,N',N'-tetramethyl ethylene diamine
$Tet^{r/s}$	tetracycline resistant (r) or sensitive (s)
Tris	tris (hydroxymethyl) aminomethane
UV	ultraviolet
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 new sigma transcription initiation factors in *Escherichia coli*, two unknown sigma cross-reacting proteins (SCRP-23 and SCRP-27A) were identified by cross-reaction with antibodies raised against region 2.2 of sigma 70. This thesis describes the mapping, sequencing and characterisation of the corresponding genes.

The gene encoding SCRP-23 was located near 652kb on the physical map of the *E. coli* chromosome. Its sequence and that of part of a downstream open reading frame were determined, and found to be closely similar to the *ahpC* and *F* genes (respectively) of *Salmonella typhimurium*. These encode the C22 and F52a subunits of the anti-oxidant enzyme, alkylhydroperoxide reductase. The identity of the *E. coli* genes was further established by their ability when introduced on plasmids into an *ahp* deletion strain to restore cumene hydroperoxide resistance. Transcription of *ahpCF* was found to be driven by two promoters: $ahpP_1$ is dependent on activation by the OxyR transcriptional regulator, whilst $ahpP_2$ is independent of this factor. Indeed $ahpP_2$ is located within the OxyR target site, and is repressed when OxyR is activated by oxidation. Thus contrary to published work, it seems that OxyR only binds significantly to its target sequence when activated. Gene expression studies indicated that C22 is produced in 9 to 12-fold molar excess over F52a, suggesting that they may not function as a simple oligomeric complex. Comparative amino acid (aa) sequence analyses identified a novel and widespread "clan" of putative antioxidant proteins related to C22, which share four regions of strong sequence similarity. Similarly, F52a was found to contain a previously unreported potentital N-terminal membrane binding domain.

The gene encoding SCRP-27A, *scrP*, was located near 3416 kb on the physical map, just 250 bp downstream of *arcB*. *arcB* encodes a transmembrane sensor-regulator of respiratory functions. Transcription analyses and protein production studies suggested that *arcB* and *scrP* may be obligatorily co-transcribed, raising the possibility that *scrP* could participate in regulation of the Arc modulon. However, in preliminary experiments a strain having *scrP* disrupted by a *kanR* insertion showed none of the increased sensitivity to redox dyes characteristic of Arc^- strains. Moreover the *scrP* :: *kanR* mutation did not affect growth of *E. coli* under a variety of conditions. The function of *scrP* remains unclear, but as in the case of SCRP-23 (AhpC) aa sequence comparisons suggest that SCRP-27A is very unlikely to be a sigma protein.

CONTENTS

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

THIS THESIS

The initial aim of this work was the identification of genes encoding novel sigma sub-units of *E. coli* RNA polymerase. This formed part of an ongoing collaboration with Ishihama and co-workers who had previously identified *E. coli* proteins which cross-reacted with antisera raised against a portion of sigma 70 which is conserved in all known proteins of the sigma 70 family (Fujita *et al* 1987). My part in this work was the characterisation of the genes encoding two such proteins, termed sigma cross-reacting proteins (SCRPS), of 23kDa and 27kDa respectively. The first part of Chapter 1 therefore describes our current knowledge of the structure and function of the conserved regions of the sigma 70 family followed by a brief discussion of each of the known alternative sigmas of *E. coli*.

The second part of Chapter 1 describes the genetics of bacterial responses to oxidative stress, concentrating on the response of *E. coli* and *S. typhimurium* to peroxide induced stress. This reflects the greater part of my work concerning the characterisation of the 23kDa SCRP, which I identified as the C22 sub-unit of alkylhydroperoxidase reductase - a peroxide stress enzyme which is elevated through the action of the OxyR transcription regulator upon exposure of *E. coli* to hydrogen peroxide.

CHAPTER 1

SEQUENCE CONSERVATION AND FUNCTION OF EUBACTERIAL SIGMAS

1.1 Introduction

E. coli RNA polymerase (RNAP) can be physically separated into two functional forms: core enzyme which is responsible for RNA polymerisation and holoenzyme which is responsible for specific initiation of transcription at distinct promoter sequences (Burgess *et al* 1969). Holoenzyme is formed by the interaction of core enzyme with a sigma subunit which determines the transcriptional specificity of RNA polymerase.

All eubacteria studied so far contain multiple sigma species and *E. coli* is no exception. So far six *E. coli* sigmas have been identified, each conferring a distinct transcriptional specificity on RNAP by altering its promoter recognition properties. The major *E. coli* sigma - sigma 70 - encoded by the *rpoD* gene is responsible for the majority of transcription whilst at least 5 minor species transcribe co-ordinately regulated gene sets in response to various stress conditions (Table 1.1).

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 holoenzyme, and promote transcription in a catalytic manner; one sigma allowing initiation by several core polymerase molecules (reviewed by Burgess and Travers, 1970).

Sigmas are homologous proteins which can be divided broadly speaking into two families, related in structure and function to sigma 70 and sigma 54 of *E. coli* respectively. As the initial aim of my work was the characterisation of genes which, it was hoped, might encode novel sigmas of the sigma 70 family, I have concentrated mostly on structure and function in that family. A much briefer discussion of the sigma 54 family is also included.

Table 1.1 *E. coli* Sigmas

Sigma	Size (kDa)	Gene	Function
70	70	*rpoD*	Major sigma factor.
S	38	*rpoS*	Major stationary phase transcription
32	32	*rpoH*	Heatshock.
E	24	not known	Extreme heatshock.
F	28	*rpoF*	Class III flagellar gene transcription.
54	54	*rpoN*	Transcription of nitrogen assimilation genes, anaerobic dehydrogenase genes and phage shock genes.

1.2 *E. coli* RNA Polymerase: A Brief Overview

E. coli RNA polymerase is a multimeric, zinc containing, DNA dependent RNA polymerase composed of at least 4 major subunits: 2α 1β $1\beta'$ 1 sigma (table 1.2). The individual subunits show no biological activity as isolated, but instead must assemble in a defined order which has been determined both *in*

vitro and *in vivo* by Ishihama and co-workers (Yura and Ishihama 1979) and can be summarised as follows:

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

Core enzyme is capable of ribonucleotide polymerisation and termination at some sites, but is unable to initiate transcription specifically. A functional sigma subunit is absolutely required to allow transcription initiation by holoenzyme (Burgess 1969).

The inactivity of isolated individual subunits of RNAP has made it difficult to ascertain their functions. However, genetic and biochemical studies have allowed some progress which is briefly summarised below.

Dimerisation of the alpha subunit is a critical initial step in RNA polymerase assembly. A substantial portion of the C-terminus of alpha can be deleted without affecting its assembly into core, whilst point mutations within the N—terminal region block RNA polymerase assembly at an early stage suggesting that the N-terminus of alpha is important for RNA polymerase assembly (Igarashi *et al* 1990, 1991, Hayward *et al* 1991). Similar studies using *rpoA* point mutants and C-terminally truncated alpha subunits *in vitro* have implicated the C—terminus of alpha in contacting type I transcription activator proteins which promote transcription by binding DNA sequences upstream of the -35 promoter element (reviewed by

Ishihama 1992).

The beta subunit is the target of the inhibitors of transcription rifampicin and streptolydigin which respectively abort initiation and block elongation (Oen and Wu 1978, Iwakuri *et al* 1973). A point mutation in the *rpoB* gene also blocks ribonucleotide polymerisation (Kashlev *et al* 1990) whilst the high conservation of sequences within the beta subunit, throughout prokaryotes and eukaryotic RNA polymerases (Rowland and Glass 1990), suggests that it may contain the catalytic centre for RNA chain polymerisation. Regions of beta have also been genetically implicated in transcription termination, proof reading and RNAP assembly (reviewed by Glass and Hayward 1993).

The beta' subunit is encoded by the same operon as beta. It is the most basic subunit and will bind DNA and other polyanions non-specifically. It is probably responsible for much of the non-specific affinity for DNA shown by core enzyme (Zillig *et al* 1970). Little else is known about beta' except that it may participate in transcription termination (Ito *et al* 1991).

The role of Sigma is discussed in detail below.

TABLE 1.2 ***E. coli* RNA Polymerase Major Subunits**

Subunit	Size (kDa)[a]	Gene[b]	Reference[c]
α	36.5	rpoA	Post and Nomura (1979)
β	150.6	rpoB	Ovchinnikov *et al* (1984)
β'	155.2	rpoC	Ovchinnikov *et al* (1982)
σ	70.3	rpoD	Burton *et al* (1981)

a) Derived from DNA sequence data presented in references (c).
b) Nomenclature of Hayward and Scaife (1976).
c) Sequence references.

1.3 The Sigma 70 Family: Functions in Brief

More than 30 sigmas of the sigma 70 family have been isolated and sequenced from a variety of eubacteria (reviewed by Lonetto *et al* 1992). Sigmas of this type promote recognition of bipartite promoters which consist of short "consensus" sequences centred near bp -35 and -10, respectively, relative to the start site of transcription (Table 1.3). The spacer region between the -10 and -35 sequences is of tightly defined length, eg usually 16-18 bp in the case of sigma 70, but of variable sequence (Helmann and Chamberlin 1988). It is widely accepted that the the members of the sigma 70 family are sequence specific DNA binding proteins, despite the inability of several groups to find any evidence that free sigma will bind to DNA (specifically or non-specifically) (Burgess *et al* 1969, Wellman and Meares 1991). However, sigma is the only subunit of RNAP holoenzyme that can be cross-linked to both the -10 and -35 promoter elements (Chenchick *et al* 1981) whilst genetic experiments have implicated two regions of sigma in interaction with the -35 and -10 promoter sequences respectively (see Sections 1.6.7 and 1.6.9 below). More recently Dombroski *et al* (1992) found that sigma 70 truncated at its N-terminus is able to bind promoters specifically, suggesting that regions of sigma involved in promoter recognition might be masked by the N-terminus of the free protein and are only uncovered on formation of holoenzyme. The evidence supporting this is discussed in Section 1.6.1 with reference to specific conserved regions of sigma.

Sigmas also play a critical role in open complex formation as indicated by the inability of core enzyme to catalyse DNA strand separation on covalently closed circular DNA (Saucier and Wang 1972). Structural and kinetic studies have shown that the opening of DNA strands at both sigma 70- and sigma 32-dependent promoters involves at least two intermediate RNAP holoenzyme-DNA complexes which show different protection patterns in DNase I, hydroxyl radical and permanganate footprinting (reviewed by Gross *et al* 1991). 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 bp -11 and +3 relative to the transcription start site (+1). During footprinting with potassium permanganate (an agent specific for single stranded DNA) the non-template strand appears to be protected and thus bound by RNAP holoenzyme (Chan *et al* 1990). In agreement with this sigma 70 can be cross-linked to the non-template strand in the unwound region of open complexes (Simpson 1987). Possible roles for regions of sigma in open complex formation are discussed in Section 1.6.8.

Table 1.3 Promoters Recognised by Bacterial Sigmas

Sigma	Organism	Consensus Sequence -35	-10
sigma 70	*E. coli*	TTGACA	TATAAT
sigma 42	*B. subtilis*	TTGACA	TATAAT
sigma 32	*E. coli*	TCTC-CCCTTGAA	CCCCAT-TA
sigma F	*E. coli*	CTAAA	CCGATAT
sigma D	*B. subtilis*	CTAAA	CCGATAT
sigma E	*B. subtilis*	GAA-AA-T	CATATT
sigma G	*B. subtilis*	TGAATA	CATACTA
Sigma	**Organism**	**Consensus Sequence -24**	**-12**
sigma 54	*E. coli*, *K. pneumoniae*, *S. typhimurium*, *Rhizobium*	CTGG-A	TTGCA

Sigma 70 and sigma 32 promoter sequences were reviewed by Hawley and McClure (1983) and Gross *et al* (1990) respectively. *B. subtilus* promoter sequences were taken from Moran (1989) and Gross *et al* (1991).

Sigma 54 promoter sequences were taken from Gussine *et al* (1986) and sigma F and sigma D promoters from Helmann (1991).

1.4 The Sigma 70 Family: Sequence Relationships

The first comparative analysis of sigmas by Landick *et al* (1984) showed that *E. coli* sigma 70 and sigma 32 (the regulator of heat-shock genes) are homologous. Further comparative studies using much larger sets of sigmas from several eubacteria confirmed that sigmas are homologous proteins (Stragier *et al* 1985, Gribskov and Burgess 1986, Hellmann and Chamberlin 1988, Lonetto *et al* 1992). Four major regions of homology have been identified, three of which are at least partially present in all primary and alternative sigmas which have been sequenced to date (figure 1.1). Gribskov and Burgess (1986) subdivided region 1 into sub-regions 1A and 1B, which are now referred to as 1.1 and 1.2 respectively (Lonetto *et al* 1992). Region 2 has been split into four sub-regions - 2.1 through 2.4 (Helmann and Chamberlin 1988) whilst regions 3 and 4 are now each subdivided into two sub-regions (Lonetto *et al* 1992, Helmann and Chamberlin 1988).

1.5 Group 1 Sigmas

Using pairwise sequence alignments and functional criteria, Lonetto *et al* (1992) divided the sigma 70 family into three groups. Group 1 sigmas are the primary sigmas of diverse organisms. They are highly homologous, sharing pairwise identities greater than 50% and are responsible for the majority of gene transcription during exponential growth. Three regions of amino acid sequence are almost invariant

within the structure of primary sigmas, and can be used to distinguish them from Group 3 alternative sigmas (discussed later). Two of these regions, the "*rpoD* box" (Tanaka *et al* 1988) and an invariant 20aa sequence, overlap the -10 and -35 promoter recognition determinants. This emphasises the close functional relationship of the primary sigmas, which recognise promoters with -35 (TTGACA) and -10 (TATAAT) consensus sequences identical to those required by the *E. coli* sigma 70.

Lonetto *et al* (1992) suggested that the small sequence differences between Group 1 sigmas (mostly conservative amino acid substitutions) are the product of genetic drift. They supported this argument by reference to their observation that the degree of sequence conservation between primary sigmas correlates with their evolutionary relationships.

Primary sigmas of Gram negative bacteria have a non-conserved, 245 aa insertion between sub-regions 1.2 and 2.1, not present within Gram positive sigmas. Recently, Hayward *et al* (1992) demonstrated that the 245 aa insertion can be deleted from *E. coli* sigma 70 without affecting its basic functions *in vivo*.

1.5.2 Group 2 Sigmas

Group 2 sigmas include *E. coli* sigma S and 3 alternative sigmas of *Streptomyces coelicolor* (HrdA, HrdC and HrdD). They show strong homology to Group 1 but are dissimilar to those primary sigmas in that they lack region 1.1. Group 2 sigmas are further distinguished

by their dispensibility during exponential growth.

The strongest regions of similarity between Group 1 and 2 sigmas include the promoter binding regions 2.4 and 4.2, which may reflect overlapping promoter specificities (Lonetto *et al* 1992). Recently Tanaka *et al* (1993) demonstrated that sigma 70 and sigma S of *E. coli* recognise a common subset of *E. coli* promoters, whilst each was found to display a distinct specificity for promoters not recognized by the other. Espinosa-Urgel and Tormo (1993) suggested that the unique promoter specificity of sigma S (which recognizes promoters of no clear consensus sequence) may be mediated by DNA bending. They based this proposition on the finding that many sigma S dependent promoters are located within intrinsically curved DNA sequences.

1.5.3 Group 3 Sigmas

Group 3 sigmas are alternative Sigmas from diverse organisms which show significant divergence from the primary sigmas. The maximal identity between the two groups is ~27%. Group 3 sigmas tend to cluster into functional sub-groups, including proteins from diverse organisms, which regulate similar physiological stress responses. Accordingly the heat shock sigmas of *E. coli* and *Citrobacter freundii* share ~94% identity, which is much higher than the 24% identity between sigma 32

and sigma 70. Similar functional clustering is observed for sigmas regulating flageller biosynthesis in enteric bacteria and *B. subtilis*, which share similar promoter specificities, (Helmann 1991) whilst *Bacillus* sporulation factors also cluster into divergent functional sub-groups.

Figure 1.1

Conserved sequences in the sigma 70 family (after Lonetto *et al* 1992). Sigma 70 and sigma 32 of *E. coli* are representative of major and alternative 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 sigma 70; evidence is for either sigma 70 or other members of its family.

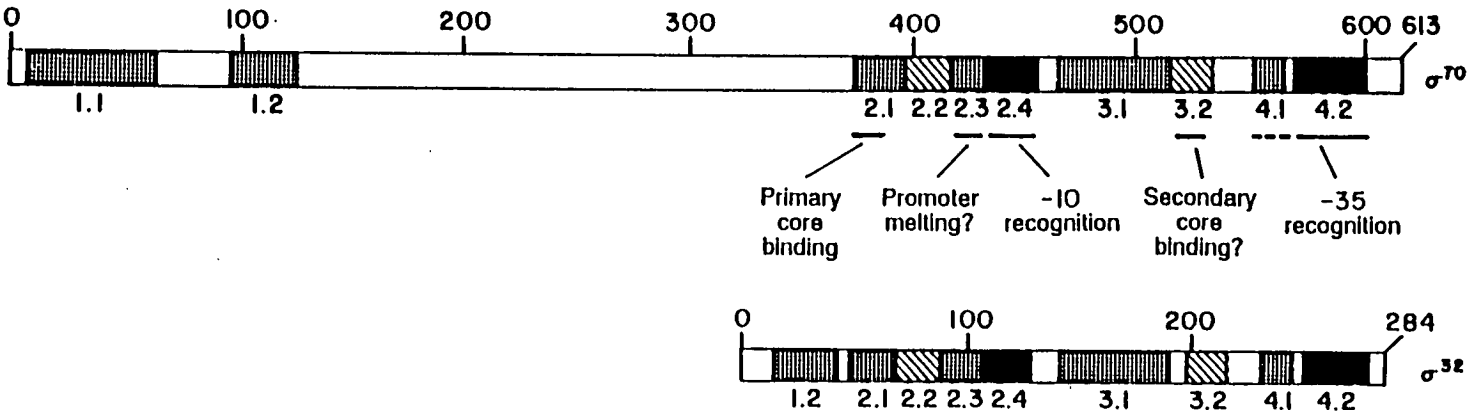

0
100
200
300
400
500
600
613
σ70
1.1
1.2
2.1
2.2
2.3
2.4
3.1
3.2
4.1
4.2
Primary core binding
Promoter melting?
-10 recognition
Secondary core binding?
-35 recognition
0
100
200
284
σ32
1.2
2.1
2.2
2.3
2.4
3.1
3.2
4.1
4.2

FIGURE 1.1

1.6 Analysis of the Conserved Regions

1.6.1 Region 1

Region 1 is sub-divided into two regions - 1.1. and 1.2. The former is present only in the primary sigmas and *E. coli* sigma S whilst region 1.2 is present (to varying degrees of similarity) within all primary sigmas and almost all alternative sigmas, an exception being *S. typhimurium* FliA (Loneto *et al* 1993). Until recently there was no evidence (and little speculation) as to the function of region 1. However, Dombroski *et al* (1992) have recently reported that deletion of the N-terminal portion of sigma 70 facilitates DNA binding by a series of truncated polypeptides fused to glutathione-S-transferase. Using a progressive set of such deletions the authors assayed both specific and non-specific DNA binding activities of sigmas deleted for all or part of region 1. The affinity of full length sigma for DNA is very low. However deletion of region 1.1 apparently bestowed non-specific DNA binding activity upon an N-terminally truncated sigma, whilst a more extensive truncation also affecting region 1.2 and sequences downstream allowed sigma 70 to specifically bind promoter DNA. The obvious interpretation of these results is that sub-region 1.1 sequesters the non-specific affinity of free Sigma 70 for DNA, whilst aa sequences between sub-regions 1.1 and 1.2 suppress

promoter recognition, either alone or in conjunction with sub-region 1.1.

The authors proposed that the N-terminus of sigma 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 enzyme in order to free the DNA binding regions of sigma. They support this with the observation that a monoclonal antibody epitope within region 1 becomes inaccessible to antibody upon RNAP holoenzyme formation. Evidence for structural changes within sigma has also been provided by Chang and Doi (1993) who noted the appearance of a distinct proteolytic fragment of *B. subtilis* sigma A, on trypsin digestion of holoenzyme, which was not generated by digestion of the free sigma. Whilst these observations might lend support to the model of Dombroski *et al* (1992) they are also consistent with a model whereby region 1 masks the DNA binding domains of free sigma 70 indirectly by promoting a conformational change within the C-terminus that renders the promoter recognition determinants inaccessible.

Finally, region 1.1 is absent from alternative sigmas leading to the prediction that some of them might bind DNA non-specifically. There is at present, no evidence to support this.

1.6.2 Region 2

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

1.6.3 Sub-Region 2.1

Sub-region 2.1 and a well conserved aa sequence immediately upstream, (aa 361-390 of sigma 70), have been deemed an essential minimal core binding region by Lesley and Burgess (1989). The authors used a set of internal deletions and frameshift mutations within *rpoD* to ascertain the boundaries of the regions of sigma that are inessential for core binding. Their results indicated that at most amino acids 361 - 390 (including the whole of region 2.1) are essential for core binding.

However, part of this region (aa 361-374) is located within the non-conserved 245 aa insertion characteristic of Gram negative primary sigmas suggesting that the essential core binding region can be further defined to the region immediately downstream of aa 374. In support of this the 245 aa insertion can be deleted from sigma 70 without affecting its basic functions *in vivo* (Hayward *et al* 1992).

1.6.4 Sub-Region 2.2

Region 2.2 has the most highly similar sequence in all eubacterial sigmas and was therefore previously proposed as a possible core binding region (Helmann and Chamberlin 1988). Region 2.2. can however be deleted without affecting the core binding activity of sigma 70 (Lesley and Burgess 1989), and at present has not been assigned any specific function.

1.6.5 Sub-Region 2.3

Helmann and Chamberlin (1988) proposed that region 2.3, in conjunction with region 2.1, might participate in DNA strand separation during open complex formation. They based this prediction upon the observed occurrence of a high proportion of aromatic amino acids within both these regions, which are reminiscent of conserved aromatic amino acids within many eukaryotic RNA and ss DNA binding proteins. Structural analyses of RNA and single stranded DNA binding proteins have shown that aromatic amino acids facilitate nucleic acid-protein interactions by stacking between the nucleotide bases (Ollis and White, 1987).

Whilst sigma 70 can be cross-linked with the non-template strand at open complexes there is little evidence at present to support a role for region 2.1 in this process. However Jones and Moran (1992) have

characterised a sigma E mutant of *B subtilis* which allows promoter recognition but not DNA strand separation by RNAP holoenzyme. The mutation (R-117-C) substitutes cysteine for arginine at position 117 within region 2.3. A similar mutation in region 2.3 of the primary *B. subtilis* sigma also inhibits DNA strand opening at RNAP holoenzyme - promoter complexes (Lonetto *et al* 1992).

1.6.6 Sub-Region 2.4

The predicted secondary structure of sub-region 2.4 is alpha helical (Helmann and Chamberlin 1988). By selecting for amino acid changes in sigma which suppress point mutations in the -10 promoter region, several groups have identified residues, all located on one face of the putative helix, which may be involved in -10 promoter recognition.

Two amino acid substitutions affecting residue 437 in region 2.4 of sigma 70 have been isolated. Q-437-H which has histidine substituted for glutamine suppresses a T-C mutation in the first position of the -10 hexamer (CATAAT) permitting a ~10 fold greater transcription from the mutant promoter than wild type sigma 70 (Waldburger *et al* 1990). A second substitution mutant, T-440-I (Siegele *et al* 1989), will suppress C or G mutations at the first position of the -10 by ~2 fold,

which is much less than that observed for Q-437-H. Both mutations described above are allele specific, that is, no other mutations in the -10 or -35 promoter regions were suppressed significantly by the Q-437-H or T-440-I mutations. Whilst T-440-I is a "loss of specificity" mutation allowing recognition on both consensus and mutant promoters, it is not clear whether Q-437-H alters the specificity of sigma 70 or results in loss of specificity for the first position of the -10, as the mutant protein cannot be assayed *in vivo* in the absence of wild type sigma 70.

Mutations analogous to Q-437-H and T-440-I have been made to similar effect in sigma A of *B. subtilis* which recognises promoters identical to sigma 70 dependent promoters (Kenney *et al* 1989).

The discovery that recognition of the same promoter base pair is altered by changes in either of two positions in region 2.4 might appear contradictory. However, crystallographic studies of known DNA binding proteins have shown that pairs of amino acid residues often co-operate in contacting a single base pair (Jordan and Pabo 1988).

Similar studies have been used to pinpoint residues in region 2.4 of an alternative sigma - sigma H of *Bacillus subtilis* which affect recognition of the first two base pairs in the -10. T-100-I suppresses a G to A mutation

at the first position of the -10 (Zuber *et al* 1989, Daniels *et al* 1990) whilst R-96-A suppresses an A to G substitution at the second position (Daniels *et al* 1990). These two mutations also cause allele specific changes of specificity, as neither mutant protein is active at wild type promoters, nor can they suppress any other promoter mutation.

An important parallel can be drawn from the results described above. sigma 70 and sigma H require quite different promoter sequences and show poor conservation in region 2.4. However, when the aa sequences are aligned, residues implicated in contacting the first position of the -10 in both proteins are adjacent, suggesting that primary and alternate sigmas may recognise promoters in a similar manner (Waldburger *et al* 1990).

1.6.7 Region 3

Region 3 is sub-divided into sub-regions 3.1 and 3.2 (Lonetto *et al* 1992). Sub-region 3.1 shows weak similarity to the H-T-H DNA binding motif but has as yet not been assigned any function. Region 3.2 has several acidic amino acids which are conserved throughout the sigma 70 family. It may play a secondary role in core binding in some sigmas as a 25 aa deletion affecting region 3.2 from sigma 32 reduced its affinity for core

enzyme (Zhou and Gross 1992). However many group 3 sigmas lack region 3.2 completely and it can be deleted from sigma 70 without significantly affecting core binding (Lesley and Burgess 1989). It seems unlikely that region 3.2 represents an essential core binding region in any sigmas.

1.6.8 Region 4

Region 4 as been divided into a pair of sub-regions - 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 region 4.1 is an amphipathic alpha helix, whilst region 4.2 shows strong characteristics of the H-T-H DNA binding motif present in many DNA binding proteins. Several lines of evidence suggest that region 4.2 recognises the -35 element of promoters.

A single amino acid substitution of alanine for valine at position 233 within region 4.2 of *B. subtilis* sigma F changes its promoter specificity to that recognised by sigma G, whilst a substitution of methionine at the same position allows sigma F to recognise sigma B dependent promoters (Margolis *et al* 1991).

Further evidence comes from the isolation of allele specific suppressor mutations in the proposed recognition helix of region 4.2 of sigma 70, which

compensate for mutations at the third and fifth positions of the -35 consensus (TTGACA). The R-584-C substitution affects the first amino acid of the proposed recognition helix and suppresses G or T substitution for C at position 5 in the -35 element (Siegele *et al* 1989). Similarly substitutions of A or T for G at position 3 in the -35 sequence of the *lac* promoter are suppressed by the R-588-H substitution at the fifth amino acid of the recognition helix (Gardella *et al* 1989).

Siegele *et al* (1989) isolated an additional set of mutations, affecting amino acids in the structural helix of region 4.2 and the amino acid sequence immediately upstream of it, which non-specifically suppress a series of promoter mutations affecting both the -35 and -10 sequences. By analogy with similar well characterised mutations affecting known H-T-H DNA binding proteins, it is likely that these sigma 70 mutations allow additional contacts with the sugar-phosphate backbone of the DNA in the vicinity of the -35 region. This could explain the **non**-specific suppression of promoter mutations by these mutant sigmas.

Little is known about the functions of sub-region 4.1. However, Lonetto *et al* (1992) suggested that it may act as a supporting structure for the helix-turn-helix DNA binding motif of sub-region 4.2. This is based on

structural similarity between region 4 and portions of the phage lambda Cro repressor and homeodomain proteins which also contain alpha helices upstream of the H-T-H motif. In these proteins the upstream helix packs against the H-T-H motif to stabilise the latter's helices.

1.7 The Sigma 54 Family

Sigma 54 proteins from at least 14 bacterial genera form a second minor class of eubacterial sigmas, which are unrelated to the sigma 70 family in structure and function. Initially identified as a positive regulator of nitrogen assimilation genes in enteric bacteria, sigma 54 has now been additionally implicated in the regulation of anaerobic dehydrogenases and phage shock operons (reviewed by Kustu *et al* 1989, Weiner *et al* 1991). Sigma 54 homologues also regulate nitrogen assimilation in *K. pneumoniae*, flagellar biosynthesis in *Caulobacter crescentus*, and toluene and xylene catabolism in *Pseudomonas putida* and are required for synthesis of pili in *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* (reviewed by Kustu *et al* 1989).

At least 14 sigma 54 homologues have been sequenced and all conform to the basic 3 domain structure (figure 1.2) first proposed by Merrick *et al* (1987). A poorly conserved acidic domain (region II) of variable length (25 - 100 residues) is flanked by a N-terminal glutamine-rich domain (region I) and a highly conserved C-terminal region (region III) ~300 amino

acids long, containing a putative H-T-H motif near the C-terminus which precedes an almost invariant block of 8 amino acids. A second putative H-T-H motif is located nearer the C-terminus but its importance has been discounted, based on statistical analysis and its poor conservation amongst the sigma 54 family members (Merrick and Chambers 1992). A pair of heptad hydrophobic repeats flanking region II have been proposed to form the elements of an intramolecular leucine zipper DNA binding domain, based on similarities between these sequences and leucine zippers in eukaryotic enhancer binding proteins (Sasse-Dwight and Gralla 1990). However, the motifs are less well conserved than in known leucine zippers, and lack adjacent basic amino acid regions which are required for DNA binding by leucine zipper proteins (Merrick and Chambers 1992). Finally according to the above prediction, sigma 54 would be the first example of an intramolecular leucine zipper protein. All other zipper domains so far identified are intermolecular.

Sigma 54 dependent promoters are markedly different in structure from those recognised by the sigma 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).

The requirements for open complex formation at sigma 54 dependent promoters are also distinct. Footprinting studies both *in vitro* and *in vivo* have failed to detect any DNA strand separation at sigma 54-containing holoenzyme-promoter

Figure 1.2

Conserved sequences in the sigma 54 family. Regions I and III are shown as open rectangles whilst region II is depicted as a solid line. Two putative H-T-H motifs and a conserved stretch of 8 aa are depicted by shaded rectangles and a black rectangle respectively, and identified below. A pair of putative leucine zipper motifs are also shown.

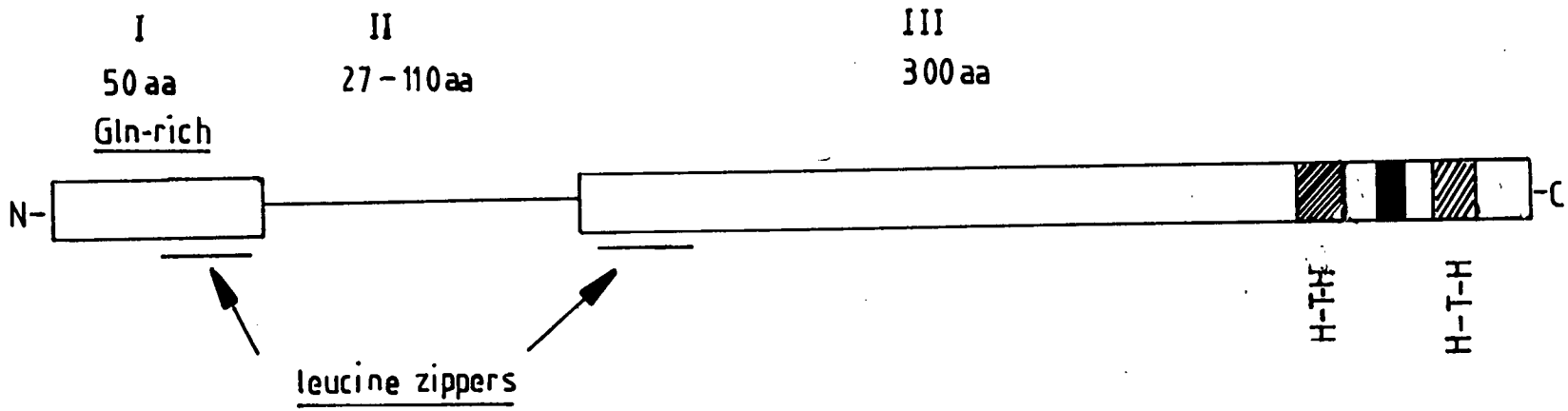

I
50 aa
Gln-rich
II
27 – 110 aa
III
300 aa
N-
-C
H-T-H
H-T-H
leucine zippers

FIGURE 1.2

complexes. Indeed, sigma 54 alone is able to bind some promoters, making contact with the DNA in a similar pattern to that observed with RNAP containing sigma 54 (Buck and Cannon 1992). This suggests that the major function of sigma 54 is specific promoter binding, without any progression to promoter melting.

The transition between closed and open complexes at sigma 54 dependent promoters is wholly dependent on additional transcriptional activator proteins. These bind to specific DNA sites which may be located between 86 bp and several kbp from the promoter. Moreoever they hydrolyse ATP to catalyse DNA strand separation (Sasse-Dwight and Gralla 1988). A requirement for contact between RNAP and activator proteins is likely as judged by electron micrographic studies, which appear to show complexes between RNAP containing sigma 54 and NTRC, the transcription activator of nitrogen regulated genes (Su *et al* 1990). Moreover, the promoter and activator binding sites need not be located at the same DNA molecule, provided the two DNAs involved are concatenatated which presumably increases the frequency of interaction between RNAP and activator (Wedel *et al* 1990).

Detailed structure/function analysis of sigma 54 is at an early stage and is subject to some controversy. Based on in-frame deletions which affected the proposed leucine zipper motifs at the N-terminus of the protein and also the disputed C—terminal H-T-H motif, Sasse-Dwight and Gralla suggested that

the proposed leucine zipper domain contacts the -12 region of the promoter and interacts with transcription activators, whilst the putative H-T-H at the C-terminus contacts the -24 region. The acidic region II was similarly implicated in DNA strand opening at the promoter.

Merrick and Chambers (1992) recently isolated a mutation in sigma 54 which causes loss of specificity for the -13 position of the *glnA*p2 promoter -12 recognition element. The mutation (substitution of lysine for arginine at position 383) is located within the strong putative H-T-H motif, as opposed to the poorly conserved more C-terminal element deleted by Sasse-Dwight and Gralla (1990). As the R-383-K substitution only affects the -13 position of the promoter, it is likely that the upstream H-T-H motif makes contact with the -12 promoter element. Merrick and Chambers argue that the observations of Sasse-Dwight and Gralla (1990) may have arisen from structural rearrangement of sigma 54 which indirectly affected recognition of the -12 promoter region.

1.8

1.8.1 *E. Coli* Alternative Sigmas

E. coli has at least 5 alternative sigmas which serve to control the expression of co-ordinately regulated gene sets required for physiological adaptation to various environmental conditions. Most alternative sigmas participate actively in regulating gene

expression, that is, the level and activity of each sigma within the cell directly affects the level of expression of its dependent genes.

The major *E. coli* sigma (sigma 70) is not considered here, as most of its known functions have been discussed in the preceding sections.

1.8.2 Sigma S

During conditions of starvation, *E. coli* cells undergo a series of metabolic and morphological changes, termed stationary phase, which facilitate their survival during starvation and confer resistance to a variety of environmental stresses including heat, oxidative agents and osmotic shock (reviewed by Matin 1991, Hengge-Aronis 1993). General gene expression is reduced during stationary phase in line with lowered metabolic rate. However, the expression of a large set of genes is increased upon starvation. Many of these are responsible for the dramatic changes in DNA topology, cell envelope structure and other aspects of physiology associated with stationary phase cells (reviewed by Matin 1991).

At least 30 such stationary phase genes are regulated by RNAP containing a novel 38 kDa sigma named sigma S Hengge-Aronis 1993). The gene encoding it -*rpoS*- was originally identified as a positive regulator of *katE*,

which encodes a stationary phase catalase HPII. It has since been identified as allelic to *nur*, a locus of mutations causing sensitivity to Near Ultra Violet Irradiation (NUV); *appR*, which is required for high level expression of *appA*-encoded acid phosphatase and *csi::lacZ* a stationary phase induced *lacZ* gene fusion (Hengge-Aronis 1993). Knock out mutations in *rpoS* reduce resistance to starvation, hydrogen peroxide (H_2O_2) and thermal stress and drastically alter the pattern of protein synthesis upon starvation (McCann *et al* 1991).

Sigma S is a group 2 sigma (Lonetto *et al* 1992) and therefore very similar to sigma 70 in regions 2 and 4, including the DNA binding regions. Recently Tanaka *et al* (1993) demonstrated that sigma S can recognise a small subset of *rpoD* dependent promoters *in vitro*, including *lacUV5*, *trp*, *rnaI* and *dnaQp2*. However, the promoter specificities of sigma S and sigma 70 are clearly different, as each will activate transcription from certain promoters not recognised by the other. However, alignment of several sigma S-dependent promoters has failed to uncover a good consensus sequence, a point emphasised by the observation that sigmas S and 70 can distinguish promoters of almost identical sequences in the -35 and -10 regions (Espinos-Urgel and Tormo 1993). Hengge-Aronis (1993) proposed that additional regulatory proteins might

modulate the promoter specificity of sigma S whilst Espinosa-Urgel and Tormo (1993) have recently demonstrated that several sigma S dependent promoters are located within intrinsically curved DNA sequences. Their results imply that DNA curvature is an important determinant in sigma S promoter recognition, perhaps preferentially aiding DNA binding by RNAP containing sigma S.

Regulation of *rpoS* is not yet well understood, but is likely to be post-transcriptional as no increase in *rpoS* mRNA has been detected upon entry into stationary phase (Lange and Hengge-Aronis 1991), whereas sigma S protein levels increase by ~7 fold (Tanaka *et al* 1993).

Regulation of sigma S activity may be an important factor in the regulation of stationary phase gene expression, although this has not been investigated.

1.8.3 Sigma F and Flagellar Biosynthesis

Helmann and Chamberlin (1987) proposed that a novel sigma might be required for transcription of flagellar and chemotaxis genes in *E. coli*, based on the observation that many such operons were preceded by sequences resembling SigD dependent promoters. SigD regulates the expression of flagellar genes in *B. subtilis*. Subsequently Arnosti and Chamberlin (1989) isolated RNAP holoenzyme containing a novel 28

kDa sigma, required for the transcription of some *E. coli* flagellar and chemotaxis genes, which they named sigma F. Sigma F is the same size as *sigD* and the two sigmas are 37% homologous (Ohnishi *et al* 1990). Indeed the cloned *sigD* gene can restore most *rpoF* regulated functions (except chemotaxis) to a sigma F deficient *E. coli* strain (reviewed by Helmann 1991).

Sigma F contributes to a regulatory cascade of gene expression which ensures the efficient regulation 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 class I genes (2 in total) express a pair of transcriptional activator proteins - FhlC and FhlD upon activation by cyclic AMP and the cyclic AMP-receptor protein (CRP). *rpoF* is one of 28 class II genes (which include structural genes for the base and hook of the flagellum) dependent on activation by FhlC and FhlD. It is required for transcription of most of the 18 class III structural genes, including *fhlC* which encodes the flagellin sub-unit protein.

Interestingly all class II genes must normally be expressed to allow class III gene expression. This regulation, which renders class III gene expression dependent on the assembly of a complete base-hook structure, may be enacted at the level of sigma F

activity. A negative regulatory locus, *flgR*, has been identified in *S. typhimurium*, which may block sigma 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 sigma 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) has suggested that these promoters may facilitate sigma 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, which is reminiscent of some positively activated sigma 70 dependent promoters, suggesting that other positive factors may be involved in late class II gene expression.

1.8.4 Sigma 32 and Heatshock

Upon exposure to increased temperature, ethanol or (to a lesser extent) phage infection and oxidative stress, *E. coli* cells mount a homeostatic response involving the elevated expression of more than 20 proteins that confer resistance to thermal stress (Niedhardt and Van Bogelen 1987). Major heat-shock proteins (HSPs) include the

molecular chaperones DnaJ, DnaK, GrpE, GroEL and GroES, which facilitate the folding and unfolding of nascent and denatured proteins and degradation or repair of aberrant or damaged proteins (Hendrick and Hartl, 1993; Georgopoulos and Welch, 1993).

Upon temperature upshift from 30°C-42°C rapid synthesis of HSPs occurs, peaking ~5-10 minutes post induction before adapting to a new (lower) steady state level, ~2 fold higher than basal levels at 30°C. The overall observed increase in HSP synthesis is ~10 fold if the temperature is upshifted to 46°C (reviewed by Bukae 1993).

Both basal level expression and induction of HSPs are dependent on RNAP holoenzyme containing the heat-shock sigma subunit, sigma 32 (Zhou *et al* 1980). Sigma 32 is a group 3 alternative sigma, by the classification of Lonetto *et al* (1992), and is much smaller than sigma 70, lacking region 1.1 and showing less overall sequence similarity in regions 2, 3 and 4. The -10 and -35 recognition elements of sigma 32 dependent promoters also differ in length and sequence when compared to sigma 70 dependent promoters (see Table 1.3). Despite these differences, biochemical and kinetic studies suggest that RNA polymerase interacts with sigma 70 and sigma 32 dependent promoters in much the same way. In particular, initiation at heat-shock promoters is not

dependent on high temperatures (reviewed by Bukae 1993).

The abundance of sigma 32 is the limiting factor in the level of HSP expression. It is present at ~10-30 copies per cell at 30°C, but accumulates rapidly by ~11 fold upon upshift to 42°C, before declining to a new steady state level ~6 fold higher than basal amounts (Straus *et al* 1987, Craig and Gross 1991). The adjustment in sigma 32 concentrations during heat-shock correlates well with induction and readjustment of HSP levels.

Transcription of *rpoH*, the sigma 32 gene, is initiated from at least 4 promoters. Of these P1 and P4 account for ~90% of *rpoH* transcription at 30°C by RNAP containing sigma 70 (Erickson *et al* 1987, Nagai *et al* 1990). A third sigma 70-dependent promoter - P5 - requires CRP-cAMP activation and is responsible for the induction of *rpoH* by ~3 fold in glucose free media, and may reflect a requirement for HSPs under starvation conditions (Nagai *et al* 1990). Transcription from P1 and P4 declines upon temperature upshift from 30°C-42°C and stops completely at 50°C, whilst transcription from P3 (which is sigma E dependent: see Section 1.8.5 below), increases to a maximum at 50°C, when it is the only active *rpoH* promoter. The function of P3 therefore is presumably to ensure production of sigma 32 at high temperatures, when sigma 70 appears to be non-functional (Wang and Kaguni 1989, Erickson and

Gross 1989). A fifth promoter, P2, is observed only in some strains of *E. coli* (Erickson *et al* 1987).

Despite the apparent complexity of *rpoH* transcription, sigma 32 regulation is primarily at the levels of protein synthesis and functional stability. Transcription of *rpoH* increases only 2 fold upon heat induction and is not responsible for adaptation at 42°C (Erickson *et al* 1987, Straus *et al* 1987). Instead a pair of cis-acting elements within the *rpoH* mRNA sequence exert a major effect on sigma 32 translation. A positive element, incorporating the initiation codon plus 20 downstream nucleotides, is required for high level expression of *rpoH* upon temperature upshift. A negative element is located between positions +110 and +247 within the *rpoH* transcript, and probably base-pairs with the positive element to sequester the translational start site and repress sigma 32 synthesis at low temperatures (Nagai *et al* 1991). A third element located between +364 and +433 has recently been implicated in the reduction of sigma 32 synthesis during adaptation to steady state levels at 42°C (Bukae 1993).

Further regulation of sigma 32 function is achieved through degradation of the protein. At 30°C it has a very short half life (~1 minute) but it is transiently stabilised ~8 fold upon temperature upshift (Straus *et*

al 1987). Sigma 32 activity is also modulated in a more subtle way. The latter conclusion is based on experiments which found that upon temperature downshift sigma 32 protein levels dropped only two fold, whilst GroEL expression declined by ~20 fold (Straus *et al* 1989). Accordingly over-expression of sigma 32 only transiently induces HSP synthesis. These regulatory mechanisms would allow a rapid mobilisation of HSPs in response to sudden thermal stress.

Whilst the detailed mechanisms of sigma 32 regulation have not been fully elucidated, mutational analysis has implicated 3 major heat-shock proteins in the regulation of sigma 32 activity. DnaK DnaJ and GrpE have been shown to facilitate Sigma 32 degradation at 30°C, whilst DnaK and GrpE are sufficient for degradation of sigma 32 during adaptation to steady state levels at 42°C after heat-shock. All 3 proteins also modulate sigma 32 activity following temperature downshift from 42°C to 30°C. It appears that these HSPs modulate sigma 32 by directly binding to the heat-shock transcription apparatus, as both DnaK and DNAJ will block transcription of sigma 32 dependent promoters *in vitro*, whilst DnaK, DnaJ and GrpE will all bind to sigma 32 *in vitro* (reviewed by Bukae 1993). Georgopoulos *et al* (1990) suggested that DnaK, DnaJ and GrpE might bind to sigma 32 and release it from RNAP holoenzyme, in order to facilitate its proteolysis.

1.8.5 Sigma E

A single *rpoH* promoter - $rpoHP_3$ - is active at extreme high temperature in *E. coli* (Erickson *et al* 1987). The factor responsible for transcription of *rpoH* 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 sub-unit of 24kDa. The protein was named sigma E to denote that it is active at extreme high temperature.

Sigma E also directs transcription of the *htrA* gene *in vitro*, which encodes a periplasmic protease that is essential for cell viability at temperatures above 43.5°C (Lipinska *et al* 1989). The activity of the *htrA* promoter is independent of *rpoH* and *rpoD*, and is elevated at 50°C *in vivo* (Lipinska *et al* 1988).

So far no other sigma E dependent genes have been identified. However several heat-shock genes are known to be expressed at elevated levels upon thermal stress in a manner independent of sigma 32, and are obvious candidates for sigma E regulation.

In the case of sigma 32, the function of $rpoHP_3$ would appear to allow continued transcription of *rpoH* under conditions where none of its other promoters is active. As described earlier, sigma 32 is highly unstable and therefore must be synthesised continuously. Further

characterisation of sigma E must await the identification and cloning of its corresponding gene (*rpoE*) and identification of more sigma E-dependent genes.

1.8.6 Sigma 54

Under conditions of nitrogen limitation the expression of nitrogen assimilation genes is activated (Magasanik 1982). Transcription of these genes requires sigma 54. However the level of transcription is not determined by the level of sigma 54 but through modulation of the activity of the transcription activator NTRC. NTRC activity is controlled through phosphorylation by the NTRB sensor-regulator protein as dictated by the level of available environmental nitrogen. Active NTRC binds, for example, to two high affinity dimer binding sites near -140 and -110bp upstream of *glnF* (encoding glutamine synthase) and stimulates open complex formation by RNAP containing sigma 54 through ATP hydrolysis and contacting RNA polymerase (Reitzer and Magasanik 1986, see also Section 1.7 above). The role of sigma 54 concentration in this process is passive. It is present at a constant low level in the cell and sequesters a small number of RNA polymerase molecules which are prebound at dependent promoters, where they await activation by NTRC.

Sigma 54 is also required for the expression of the *fdhF* and *hyd* genes during anaerobiosis. These genes encode sub-units of formate hydrogen lyase, which degrades formate to CO_2 and H_2 in the absence of alternative electron acceptors such as nitrate (reviewed by Kustu *et*

al 1989). Transcription of the *fdhF* and *hyd* genes is dependent on activation by the FhlA transcription activator.

Finally, at least 4 *psp* genes, encoding phage shock proteins, which are expressed in response to infection by filamentous phages and to heat-shock and ethanol, are also transcribed by sigma 54. The level of transcription of *psp* genes is positively autoregulated by the *pspC* gene product which interestingly contains a potential leucine zipper motif characteristic of many eukaryotic activator proteins (Weiner *et al* 1991).

1.9 Oxidative Stress Responses of *E. coli* and *S. typhimurium*

1.9.1 Introduction

Aerobic organisms, which utilise molecular oxygen (O_2) as the terminal electron acceptor in oxidative phosphorylation, must deal with toxic reactive oxygen species which are by-products of aerobic metabolism. Active oxygen species damage many cellular components and have been implicated in several human degenerative disorders including rheumatoid arthritis, atherosclerosis and cancer. They are also involved in the ageing process.

Accordingly, aerobic organisms have developed complex defence systems to counteract the oxidative threat imposed by reactive oxygen. Catalases and superoxide dismutases, which destroy hydrogen peroxide and superoxide radicals respectively, have been isolated from several organisms. However, in most cases little is known about the genetic control of anti-oxidant defences.

The best studied organisms in this respect are the Gram negative enteric bacteria *E. coli* and *S. typhimurium*. Both organisms although normally inhabitants of the outer reaches of the anaerobic intestinal tract, are

able to grow aerobically and have for example been isolated from sewage and soil (Storz *et al* 1990). Consequently they must and can mount defensive responses against oxidative stress. In fact they have (owing to their well understood genetics) proved invaluable in the characterisation of oxidative stress responses.

In the following sections I will concentrate on our current knowledge of the regulation of the response to hydrogen peroxide, and will discuss in less detail the superoxide stress response and the *rpoS* regulon. I will not consider in detail the nature of active oxygen species nor the damage they cause. These are very complex topics in their own right and are not the main focus of this thesis.

1.9.2 Active Oxygen Species

Molecular oxygen (O_2) is paramagnetic, having two unpaired electrons of parallel spin in separate bonding orbitals. Consequently it will not readily undergo reduction by accepting a pair of electrons, as paired electrons in atomic and molecular bonding orbitals are most commonly of antiparallel spin. However unless O_2 does undergo reduction by accepting a pair of electrons of parallel spin, it will generate reactive oxygen species.

Of the major active oxygen species only the superoxide radical (O_2^-) hydrogen peroxide (H_2O_2) and the hydroxyl free radical ($OH^\cdot$) are considered here. The former pair elicit well characterised physiological defences (the subject of this thesis), while the latter is thought to mediate much of the damage generated by both H_2O_2 and O_2^- (reviewed by Farr and Kogoma 1991). Both of the latter species are produced as a result of aerobic metabolism.

1.9.3 Superoxide (O_2^-)

O_2^- is generated by the single electron reduction of molecular oxygen by a wide variety of cellular and exogenous agents. Metabolic O_2^- is produced primarily through the auto-oxidation of membrane associated electron-transport enzymes and, to a lesser extent, by cytoplasmic enzymes such as glutathione reductase (Imlay and Fridovich 1991, Farr and Kogoma 1991). Non-enzymatic production of O_2^- within the cell is due to the auto-oxidation of intra-cellular thiols, flavins and catechols and by redox cycling of electrophilic quinols such as ubiquinone. Exogenous sources of O_2^- are the laboratory compounds Plumbagin (PL) and Menadione (MD), which are electrophilic quinols, and the dipyridyl compound Paraquat (PQ) (reviewed by Farr and Kogoma 1991).

O_2^- is only weakly reactive in aqueous solution but nevertheless poses a substantial oxidative threat to *E. coli* cells. The bacterium possess two superoxide dismutases (SOD's) (one Fe SOD and one Mn SOD) which rapidly destroy O_2^- during aerobiosis. In contrast *sodAB* double mutants (lacking both SOD activities) accumulate O_2^- at around 5 μmol per second, and are unable to grow in minimal medium unless supplemented with branched chain, aromatic and sulphur-containing amino acids (Carloz and Touati 1986). This is probably due to the reactivity of O_2^- with $[Fe\text{-}S]_4$ clusters in biosynthetic enzymes, such as dihydroxy acid dehydratase, required for synthesis of branched chain amino acids and aconitase, a citric acid cycle enzyme, which are both rapidly inactivated by O_2^- (Gardner and Fridovich 1992).

O_2^- also causes membrane damage, as indicated by the loss of proton motive force dependent and independent membrane transport produced by superoxide radical generators. Moreover the amino acid auxotrophies of *sodAB* mutants can be partially suppressed by addition of external osmolytes to nutrient-limited media such as to balance the concentrations of intracellular and external osmolytes, presumably by preventing leakage of essential nutrients across the damaged cell membrane. Further mechanical membrane damage would also be prevented by such treatment (Gardner and Fridovich

1992).

O_2^- is also likely to promote DNA damage. *sodAB* mutants show increased rates of accumulation of mutations, which have been attributed to the action of exonuclease III on apurinic and apyrimidinic (AP) sites O_2^--damaged DNA. Such sites of base loss are premutagenic lesions, but are converted to mutagenic lesions by exonuclease III (Farr *et al* 1986, Hoerter *et al* 1989). Further lethal DNA strand breaks are likely to occur through the participation of O_2^- in a series of reactions termed Haber-Weiss chemistry which generate $OH^{\cdot}$ and perferryl radicals (FeO^{++}). These are significantly more reactive with cellular components than is O_2^-.

1.9.4 Hydrogen Peroxide (H_2O_2)

The major cellular source of H_2O_2 is the SOD catalysed dismutation of O_2^-. Spontaneous dismutation of O_2^-, NUV-irradiation and cellular oxidases also represent significant sources of H_2O_2 (Farr and Kogoma 1991). H_2O_2 is only weakly reactive towards organic molecules. It will oxidise protein thiols and reduced glutathione, and is capable of oxidising transition metals. The latter of these reactivities is the most significant, as the oxidation of Fe^{++} by H_2O_2 generates $OH^{\cdot}$ radicals which cause DNA strand breaks both *in vitro* and *in vivo* (Greenberg *et at* 1986).

Imlay and Linn (1986) found that killing of *E. coli* by H_2O_2 is bimodal. At low H_2O_2 concentrations (1-3mM) killing requires active metabolism, and is enhanced by mutations in *recA*, *xthA* and *polA*, which block recombinational DNA repair (*recA*) and repair of DNA strand breaks respectively. At higher concentrations (>20mM) lethality does not require active metabolism.

H_2O_2 also causes membrane damage, as indicated by the inhibition of proton motive force dependent and independent membrane transport upon exposure of *E. coli* to H_2O_2 (Farr *et al* 1988).

1.9.5. The Hydroxyl Radical ($OH^{\cdot}$)

$OH^{\cdot}$ radicals are produced by the photolysis of tryptophan residues, by radiolysis of H_2O and by the Fenton reaction, shown below:

$$H_2O_2 + Fe^{++} + H^{+} + e^{-} \longrightarrow OH^{\cdot} + H_2O + Fe^{+++}$$

(reviewed by Farr and Kogoma 1991). $OH^{\cdot}$ is highly reactive and will oxidise most cellular components. This high reactivity limits the diffusion distance of $OH^{\cdot}$ to a few nanometers so that cellular damage is likely to be confined to the site of $OH^{\cdot}$ generation.

As noted above, much of the damage caused by O_2^{-} and H_2O_2 may occur through the generation of $OH^{\cdot}$. It causes DNA strand breaks by fragmenting sugar residues, generating 3' blocked phosphate or phosphoglycolate termini which can potentially block DNA replication/ repair (see Section 1.12 later). $OH^{\cdot}$ may also be mutagenic as it can undergo addition reaction with deoxyguanosine to produce 8-hydroxy deoxyguanosine. It is also thought to be an instigator of lipid peroxidation (Farr and Kogoma 1991).

Lipid peroxidation, and subsequent breakdown of lipid hydroperoxides to lipid radical compounds, can result in a chain reaction which generates ~50 damaged molecules per originally peroxidised lipid (Kappus and Sies 1981). In addition to causing membrane damage the intermediates

and products of lipid peroxidation are mutagenic, alkylating DNA or generating inter or intra-strand cross links (Summerfield and Tappel 1983).

1.10 Oxidative Defence Regulons of *E. coli* and *S. typimurium*

Aerobically growing bacteria combat the threat posed by oxidative stress by mobilising complex adaptive responses which are largely specific for the insulting agent, and are subject to co-ordinated genetic regulation. *E. coli* and *S. typhimurium* both synthesise distinct sets of proteins when challenged with sublethal doses of H_2O_2 or O_2^-, which counteract the effects of these agents and confer resistance to subsequent lethal challenges. The following sections describe current knowledge of the regulation and nature of the responses to H_2O_2 and O_2^- -generated stress with particular emphasis on the former. The role of the *rpoS* regulon in combating oxidative stress is also briefly discussed.

1.10.1 The Peroxide Stress Response

Evidence that *E. coli* and *S. typhimurium* cells could mount an adaptive response when challenged with hydrogen peroxide (H_2O_2) was provided by Demple and Halbrook (1983) and Christman *et al* (1985), who demonstrated that cells pretreated with sublethal doses of H_2O_2 (5μM-50μM) could survive subsequent challenges with H_2O_2 concentrations (5mM-25mM) that were lethal to naive

characterisation of sigma E must await the identification and cloning of its corresponding gene (*rpoE*) and identification of more sigma E-dependent genes.

cells. This response may consist in part of an inducible DNA repair pathway in view of the observations that P1 or lambda bacteriophages, previously damaged *in vitro* by H_2O_2, are reactivated with ~2 fold greater efficiency in H_2O_2-adapted *E. coli* than in naive cells (Demple and Halbrook 1983, Farr *et al* 1985). However, Christman *et al* (1985) and Imlay and Linn (1986) failed to observe any such difference between the efficiency of reactivation of H_2O_2-damaged P22 or lambda phage in H_2O_2-adapted versus non-adapted *S. typhimurium* or *E. coli* cells, respectively. If any inducible DNA repair pathway plays a major role in the adaptive response, it must be distinct from the *recA* dependent SOS DNA repair pathway, as RecA deficient cells can still adapt to peroxide stress (Demple and Halbrook 1983, Christman *et al* 1985). The RecA pathway may nevertheless be induced as a secondary response to oxidative DNA damage, because *recA* mutants show increased sensitivity to H_2O_2 despite their ability to mount an adaptive response (Imlay and Linn 1988). Moreover strains with deletions of the *oxyR* locus, which is involved in the regulation of the peroxide stress response (see Section 1.10.2) derepress the SOS regulon at much lower concentrations of H_2O_2 than do otherwise isogenic $OxyR^+$ strains (Goerlich *et al* 1989).

The peroxide stress response can be evoked by superoxide generating drugs (Greenberg and Demple 1988) and by near

ultraviolet (NUV) radiation (Eisenstark 1989). The above agents promote the formation of intracellular H_2O_2 by their reactions with cellular components. However *E. coli* cells pretreated with sub-lethal doses of plumbagin or H_2O_2 (menadione and paraquat have not been tested) do not develop reciprocal cross resistance to the other agent (Farr *et al* 1985).

Conflicting evidence exists concerning the cross resistance of H_2O_2 and NUV stressed cells. Whilst some authors have reported reciprocal cross resistance of adapted cells to H_2O_2 and NUV, others have reported that on the contrary H_2O_2 pretreatment sensitises *E. coli* to NUV irradiation (reviewed by Eisenstark 1989).

Additionally the observed cross resistance of H_2O_2 adapted *S. typhimurium* cells to heat stress (Christman *et al* 1985) does not require the induction of the heat-shock regulon (Demple 1991). The cellular reaction to peroxide stress is therefore a distinct adaptive response specific for the oxidative challenge posed by organic hydroperoxides.

1.10.2 Peroxide Stress Proteins and the OxyR Regulon

Both Demple and Halbrook (1983) and Christman *et al* (1985) found evidence that the adaptation of *E. coli* and *S. typhimurium* to peroxide stress required *de novo* protein synthesis, in that addition of chloramphenicol

during the pretreatment of cultures with H_2O_2 prevented them from developing the ability to survive subsequent challenges with higher doses of peroxide. Consistent with this, a distinct new pattern of protein synthesis is observed when *E. coli* and *S. typhimurium* strains are treated with low concentrations (30 μM) of H_2O_2.

Around 30 *S. typhimurium* proteins and between 30 and 40 *E. coli* proteins are expressed at elevated levels post-induction (Christman *et al* 1985, Van Bogelen *et al* 1987, Greenberg and Demple 1988, Walkup and Kogoma 1989). At least 12 so-called "early" proteins are synthesized at elevated levels only in the first 30 minutes of induction, whilst a further ~18 "late" proteins continue to be synthesized for at least 60 minutes after addition of H_2O_2. These "peroxide stress proteins", only a few of which have been identified, include anti-oxidant enzymes such as HPI catalase, alkylhydroperoxide reductase and a few heat shock proteins, and are listed in Table 1.4

Nine of the twelve early peroxide stress proteins are constitutively over-expressed in certain *E. coli* and *S. typhimurium* mutants which were selected by Christman *et al* (1985) as showing increased resistance to H_2O_2. This constitutive phenotype was found to be caused by mutations in a distinct genetic locus, termed *oxyR*, (Christman *et al* 1985) which was shown to be linked to

argH (marked by Tn*5* insertions) near 89.5 minutes on the standard *E. coli* genetic map (Bachman 1990). Subsequently deletion of *oxyR* was mediated by Tn*10* insertions in *argH* and *btuB* (located at 89.5 and 89.7 minutes respectively on the genetic map). This mapped the *oxyR* locus between these two markers, near 89.6 minutes, owing to the unidirectional nature of the Tn*10* mediated deletions. Christman *et al* (1985) mapped the *S. typhimurium oxyR* locus to a similar relative position. As *oxyR* deletions abolished H_2O_2 induction of the 9 early proteins, the authors concluded that OxyR exerts a positive effect which is essential for the induction of the OxyR regulon.

Identified proteins of the *S. typhimurium* OxyR regulon include 2 electromorphs of HP1 catalase, the C22 and F52a subunits of alkylhydroperoxide reductase (Ahp) and 2 heat shock proteins (Table 1.4). Consistent with this, the activities of HP1 catalase and Ahp were found to be elevated by 50 fold and 5-20 fold respectively in OxyR constitutive mutants. Similarly, glutathione reductase and Mn-containing superoxide dismutase (MnSOD) activities were increased by ~4 fold each. However, neither activity was correlated with any of the 9 early peroxide stress proteins (Christman *et al* 1985).

In the same study, a similar analysis of *E. coli* proteins identified 4 of the OxyR regulated proteins as

two electromorphs of HPI catalase and the two subunits of Ahp. However, none of the *E. coli* heat shock proteins were found to be present at elevated levels in OxyR constitutive ($OxyR^{(cons)}$) mutants. The authors did find that, as in *S. typhimurium*, MnSOD was expressed at elevated levels in $OxyR^{(cons)}$ mutants. However, subsequent analysis of *sodA* (MnSOD) transcription in *E. coli* revealed that the concentration of *sodA* mRNA was not elevated by H_2O_2 induction (Touati 1988). In agreement with this Bowen *et al* (1988) found that H_2O_2 treatment did not lead to increased MnSOD protein synthesis but that it could be induced by paraquat in *oxyR* deletion mutants, indicating that in *E. coli*, MnSOD induction is independent of OxyR.

TABLE 1.4 GENES INDUCIBLE BY H_2O_2 in *E. coli* and *S. Typhimurium*

Gene	Organism	Gene Product	OxyR Regulation		References
			E. coli	*S. Typhimurium*	
ahpC *ahpF*	*E. coli* and *S. typhimurium*	C22 and F52a subunits of alkylhydroperoxide reductase (Ahp)	Yes	Yes	Christman *et al* (1985), Greenberg and Demple (1986) Vanbogelen *et al* (1987), Storz *et al* (1989)
katG	*E. coli* and *S. typhimurium*	HP1 catalase	Yes	Yes	Christman *et al* (1985), Greenberg and Demple (1986) Vanbogelen *et al* (1987), Walkup and Kogoma (1989).
gor	*S. typhimurium*	Glutathione reductase	No	Yes(b)	Christman *et al* (1985).
oxyS	*E. coli*	Regulatory RNA	Yes	Not Known	Tartaglia *et al* (1992)
orfO	*S. typhimurium*	Not known	Not known	Yes	Tartaglia *et al* (1992).
groES	*E. coli*	GROES±(a)	No	No	+Vanbogelen *et al* (1987), -Greenberg and Demple (1986).
groEL	*E. coli*	GROEL±(a)	No	No	+Morgan *et al* (1986), -Vanbogelen *et al* (1987).
dnaK	*E. coli* and *S. typhimurium*	DnaK	No	Yes(c)	Morgan *et al* (1986), Vanbogelen *et al* (1987).
recA	*E. coli*	RecA	No	No	Greenberg *et al* (1986), Vanbogelen *et al* (1987).
sodA	*S. typhimurium*	Manganese containing superoxide dismutase (MnSOD)	No	Yes(b)	Christman *et al* (1985).

(a) ± denotes that both positive and negative observations have been reported.
+ denotes positive observation, - denotes negative observation.
(b) Enzyme activity constitutively elevated in *S. typhimurium* OxyR2 mutant
(c) Constitutively elevated in OxyR2 mutant.
General Reference: Farr and Kogoma (1991).

1.10.3 OxyR Regulation

The *oxyR* locus has been cloned and sequenced independently by three groups. Christman *et al* (1989) and Tao *et al* (1989) cloned *oxyR* by complementation of H_2O_2 sensitivity in OxyR deletion mutants, whilst Bolker and Kahmann (1989) cloned *oxyR* as the repressor of the phage Mu *mom* gene (which encodes a DNA modification function).

All 3 groups found that the *oxyR* locus consists of a single open reading frame, encoding a 34 kDa polypeptide related to the LysR family of transcriptional activators, and sharing ~25% identity with the LysR family consensus sequence (Henikoff *et al* 1988). The strongest sequence similarities are clustered within an N-terminal region, including a helix-turn-helix DNA binding motif shared by all the LysR family proteins. The LysR family, which includes the regulators of amino acid biosynthesis LysR, MetR; IlvY and CysB, plus NodD, a regulator of root nodule formation in *Rhizobium* shares several other characteristics with OxyR, including negative autoregulation. This was shown for OxyR by a five fold increase in expression of an *oxyR:lacZ* transcriptional fusion in an *oxyR* deletion background as compared to an $OxyR^+$ strain, and by the ~8 fold down-regulation of a similar fusion consequent on over-expression of OxyR from a multicopy plasmid (Christman

et al 1989, Tao *et al* 1991). *oxyR* is also divergently transcribed from the *oxyS* gene which it regulates, as are several other members of the LysR gene family with respect to operons under their control. Moreover all LysR family members are single component activator proteins (Henikoff *et al* 1988).

As suggested by its relationship with the LysR family of transcriptional activators, OxyR regulates gene expression at the level of transcription. Quantitation of *katG* transcripts from $OxyR^{+}$ and $OxyR^{(cons)}$ backgrounds revealed a 50 fold increase in their steady state level in the $OxyR^{(cons)}$ background, which concurs with the 50 fold elevation of HPI catalase activity previously found in $OxyR^{(cons)}$ strains (Morgan *et al* 1986). Similarly Tartaglia *et al* (1989) observed a significant increase in the rate of *S. typhimurium ahpC* transcription in an $OxyR^{(cons)}$ mutant compared to a wild type strain.

DNA footprinting analysis has defined OxyR binding sites at the *katG* and *oxyR* promoters of *E. coli* and the *ahpC* promoter of *S. typhimurium* (Tartaglia *et al* 1989, Tao *et al* 1991), whilst Bolker and Kahmen (1989) used MPE-Fe(II) footprinting to define the OxyR binding site at the *mom* promoter of phage Mu.

At *PkatG*, OxyR protects a region of DNA between position -76 and -32 on the top strand, and between positions -33

and -80 on the bottom strand, relative to the start site of the *katG* transcript (Tartaglia *et al* 1989, Tao *et al* 1991). Similarly a region between -79 and -35 on the top strand of the *ahpC* promoter is protected by OxyR, against DNase I digestion (Tartaglia *et al* 1989). At both promoters the OxyR binding site overlaps the -35 consensus sequence, suggesting that OxyR activates transcription through contacting RNA polymerase.

Recently, Tao *et al* (1993) found evidence that OxyR probably makes contact with RNA polymerase at the C-terminus of the alpha subunit, as wild type RNA polymerase, but not mutant RNA polymerases containing C—terminally truncated alpha subunits, were able to activate transcription at the *katG* and *ahpC* promoters *in vitro*. This is consistent with a large body of evidence that many transcriptional activators contact RNA polymerase at the C—terminus of the alpha sub-unit (reviewed by Ishihama 1992). Tao *et al* (1993) also showed that OxyR stabilises the initial binding of the RNA polymerase at the *katG* promoter in a co-operative manner, as judged by the ability of RNA polymerase to protect the *katG* promoter from DNase I digestion <u>only</u> in the presence of OxyR.

At the *oxyR* promoter, OxyR represses transcription of its own gene by binding to a site between +15 and -35 on the coding strand, and +21 and -27 on the non-coding

strand, relative to the mRNA start site (Tao *et al* 1991, Christman *et al* 1989). Presumably, by binding over the -10 consensus sequence of the promoter, OxyR blocks RNA polymerase binding and so represses transcription or promoter-opening.

Interestingly the OxyR binding site at the *oxyR* promoter overlaps the -35 promoter element of the *oxyS* gene, encoding a small 107 nucleotide RNA molecule OxyS which is a global regulator of gene expression (Storz 1992). *In vitro* transcription assays have shown that OxyR can activate *oxyS* transcription while repressing transcription of the *oxyR* gene (Storz *et al* 1990, Tao *et al* 1993). Transcriptional repression of the phage Mu *mom* gene is achieved by OxyR binding upstream of the *mom* promoter, protecting a segment of DNA between positions -92 and -50 on the top strand, relative to the mRNA start site (Bolker and Kahmen 1989). The OxyR binding site overlaps the target of the bacteriophage Mu C protein, which is an activator of *mom* transcription. OxyR may therefore interfere with C protein binding at the *mom* promoter, or with C protein - RNA polymerase interactions.

Comparison of the nucleotide sequences of the known OxyR binding sites initially revealed no obvious sequence similarity. The very different DNase I footprints produced by OxyR binding also suggested a lack of common

structural features (Tartaglia *et al* 1990). Nevertheless OxyR is apparently able to recognise the various target sequences with a high degree of specificity, as judged by competition experiments using random DNA sequences (Tartaglia *et al* 1989 and 1992). Later, alignment of the known OxyR target sequences (plus the newly determined target sequence upstream of an unidentified open reading frame, *orf0*) revealed several positions of degenerate homology where only 2 of the 4 possible base pairs occurred within the OxyR binding sites (Tartaglia *et al* 1992). Subsequent experiments showing protection of several of these sites by OxyR, in methylation interference assays, led the authors to conclude that OxyR specificity is dependent on a multidegenerate recognition code. The authors proposed a model for recognition (based on the base pair discrimination models of Seeman *et al* 1976) whereby OxyR recognises similar patterns of hydrogen bond acceptor and donor atoms displayed by degenerate base pairs at important contact sites in the major groove of the DNA duplex.

In addition to its observed sequence specificity, OxyR is methylation sensitive. It will only repress the *mom* promoter when 3 Dam methylase targets (GATC) within its OxyR binding site are unmethylated or hemimethylated (Bolker and Kahmen 1989). The inference that OxyR might therefore regulate the transcription of oxidative

stress genes in a methylation dependent manner is supported by the finding that Dam^- mutant strains of *E. coli* are hypersensitive to H_2O_2 (Yallaly and Eisenstark 1990).

1.10.4 Direct Activation of OxyR by Oxidation

Immunological studies which directly quantified amounts OxyR'-'LacZ fusion protein, and analyses of the enzymatic activity of a similar fusion protein, revealed that OxyR synthesis is not increased at the onset of peroxide stress (Storz *et al* 1990, Tao *et al* 1991). This suggests that a change in OxyR activity may be responsible for the activation of the OxyR regulon. Consistent with this, characterisation of the OxyR2 constitutive mutant of *E. coli* revealed that its constitutive phenotype was due to a single amino acid substitution at residue 234 Ala to Val (Christman *et al* 1989).

The hypothesis that OxyR might be directly activated by oxidation was based on the observations that OxyR enriched cell extracts, or purified OxyR protein prepared from non-stressed aerobically grown cells, could activate transcription of the *ahpCF* and *katG* genes, whilst extracts of OxyR protein prepared under highly reducing conditions (100mM DTT) could not (Storz *et al* 1990). Furthermore, reversible activation or

inactivation of purified OxyR protein was simply achieved respectively by removal of DTT (by dialysis) or its addition. Further experiments also demonstrated that OxyR was inactivated by addition of catalase, or by anaerobic conditions, as judged by its ability to activate transcription *in vitro*.

Oxidation appears to induce a conformational change in the structure of OxyR, as the DNase I protection patterns displayed by the reduced form of the protein at the *katG, ahpC* and *oxyR* promoters differ from those produced by the active oxidised protein. In particular, a distinct DNase I hypersensitive site occurs near the centre of the DNA target only when it is protected by the reduced protein, at all promoters tested, whilst the extent of the footprints generated by the reduced protein at the *katG* and *oxyR* promoters also differ from those of the oxidised form (Storz *et al* 1990). The affinities of the reduced and oxidised forms of OxyR for their specific binding sites appeared not to differ significantly however, leading to the suggestion that OxyR is permanently bound at its target sequences and is induced by oxidation (due to increased flux of H_2O_2 into the cell) only to change its conformation and thereby activate transcription of the OxyR regulon genes (Storz *et al* 1990, Tartaglia *et al* 1992).

The nature of the redox active centre of OxyR has not

yet been determined. However, all but one of the six Cys residues apparent in the OxyR primary structure can be substituted without loss of function, suggesting that intramolecular disulphides bonds are not required for OxyR function. Substitution of Cys 199, however, inactivates OxyR without affecting its DNA binding function at the *oxyR* promoter (Storz *et al* 1990). Cys 199 might therefore represent a redox active thiol or might form an intermolecular disulphide bond. Multimeric forms of OxyR have not been observed in solution (Storz *et al* 1990). However, the large sizes of OxyR binding sites (~45bp) and sequence evidence for half sites within them, suggest that two (or more) molecules of OxyR may bind, but perhaps only by dimerising on interaction with the DNA duplex (Tartaglia *et al* 1992).

1.10.5 OxyR Regulon Proteins
Alkylhydroperoxide Reductase

OxyR constitutive mutants of *E. coli* and *S. typhimurium* are more resistant than wild type strains to a range of oxidative agents including alkylhydroperoxides (Christman *et al* 1985). Resistance to alkylhydroperoxides is mediated by a distinct peroxidatic activity comprising the C22 and F52a polypeptides of alkylhydroperoxide reductase (Ahp), which can use NADH or NADPH as electron donors for the direct reduction of alkylhydroperoxides to their

corresponding alcohols.

The two polypeptides of Ahp are encoded by a single locus, located between *ent*A and *lip*, near 13 minutes on the *S. typhimurium* genetic map. The *ahp* locus has been less accurately mapped to a similar location on the genetic map of *E. coli* K12 (Storz *et al* 1989). Tn*10* mediated deletions of this locus result in hypersensitivity to cumene hydroperoxide (CHP) and abolish synthesis of the C22 and F52a polypeptides, as judged by 2-dimensional electrophoresis of [^{35}S]met labelled proteins. Subsequent cloning of the *ahp* loci of *E. coli* and *S. typhimurium* was achieved by complementation of the CHP sensitivity of *ahp* deletion mutants (Storz *et al* 1990).

The DNA sequence of the *ahp* locus of *S.typhimurium* was determined by Tartaglia *et al* (1990), who also concluded that the *ahp C* and *F* genes (encoding the C22 and F52a proteins respectively) are probably co-transcribed, as a number of independently isolated Tn*10* insertions in the *ahp* locus abolished synthesis of both proteins.

Ahp was purified by Jacobson *et al* (1989) using an assay based on the reduction of a model hydroperoxide-cumene hydroperoxide-which was shown to require two separate components, identified as the C22 and F52a polypeptides. These were present at elevated levels in OxyR constitutive mutants. The reconstituted activity,

which requires NAD(P)H, is inhibited by the addition of thiol reagents, whilst both the C22 and F52a polypeptides appear to contain redox active disulphides, based on titration experiments using 5-5'-dithiobis (2-nitrobenzoic acid) (DTNB).

The absorption spectrum displayed by the F52a protein is characteristic of a flavoprotein. It has been shown to contain 1 FAD cofactor per subunit and is able to reduce a variety of electron acceptors in the presence of NAD(P)H, suggesting that it is a flavoprotein disulphide oxidoreductase (Jacobson *et al* 1989). This observation is strengthened by comparative protein sequence analysis which reveals that the F52a protein is closely related to thioredoxin reductase, a well characterised disulphide oxidoreductase. This transfers electrons from NAD(P)H via an FAD co-factor and a redox active disulphide, to directly reduce a disulphide on the small non-enzymatic protein thioredoxin - a reductant in several cellular processes (Williams 1982).

The similarity between the F52a protein and thioredoxin reductase includes perfect conservation of the redox active cysteines of thioredoxin reductase within a NAD(P) binding segment which is flanked (at a distance) by elements of an FAD cofactor binding domain. Both proteins show a lesser degree of similarity with a second, larger class of disulphide oxidoreductases

(comprised of glutathione and mercuric reductases and dihydrolipoamide dehydrogenases from a number of organisms). The similarities between the two classes correspond to their FAD and NAD binding domains. The proteins of the latter class however have a distinctive and highly conserved active site (containing a redox active disulphide) within the FAD binding domain, and are likely to have evolved separately from the proteins of the thioredoxin reductase and F52a class (Russell and Model 1988, Tartaglia *et al* 1990).

The requirement for C22 protein to confer upon F52a reactivity towards alkylhydroperoxides strengthens the comparison with the thioredoxin - thioredoxin reductase system, and allows speculation concerning the mechanism of peroxide reduction by Ahp. Jacobson *et al* (1989) proposed that the C22 protein is the site of peroxide substrate reduction coupled to NAD(P)H oxidation on the F52a subunit, in a mechanism that involves transfer of electrons from NAD(P)H to the FAD cofactor, then to a redox active disulphide on the F52a protein, before transfer to the alkylhydroperoxide substrate via a second disulphide on the C22 protein (figure 1.3). Claiborne *et at* (1992) suggested that the reduced dithiol form of the C22 protein must have unique properties which allow its direct oxidation by the alkylhydroperoxide (ROOH) substrate. The authors proposed that initial oxidation of a Cys-SH by the ROOH

substrate yields a cysteine sulphenic acid (Cys-SOH) and the R-OH alcohol product, and is rapidly followed by the nucleophilic displacement of the second Cys-SH to yield the C22 protein disulphide and H_2O. Direct chemical and spectroscopic evidence for this proposed mechanism has yet to be provided.

A wide range of alkylhydroperoxides, including linoleic hydroperoxide and thymine hydroperoxide, act as substrates of Ahp *in vitro*. This suggests that hydroperoxides of lipids and perhaps DNA, both of which are formed during intracellular free radical chain reactions of unsaturated lipids and nucleic acids, may be substrates *in vivo* (Jacobson *et al* 1989). H_2O_2 might also be a substrate for Ahp *in vivo*, as over-expression of Ahp results in increased resistance of OxyR deletion strains to H_2O_2 (Storz *et al* 1989). However, the enzyme is rapidly inactivated by H_2O_2 *in vitro*, suggesting that Ahp-mediated H_2O_2 resistance in OxyR deletion strains may be through detoxification of secondary hydroperoxides generated by the initial H_2O_2 insult.

Although the substrate(s) of Ahp *in vivo* are not yet known, there is ample evidence that the enzyme plays an important role in the prevention and reversal of oxidative cellular damage. Deletion of the *ahp* locus results in hyper-sensitivity to cumene hydroperoxide

(Storz *et al* 1989). On the other hand over-expression of Ahp, either due to *ahp* multicopy or as a result of mutations at the chromosomal *ahp* locus, suppresses the high rate of spontaneous mutation observed in OxyR deletion strains, and restores resistance to H_2O_2 and the redox cycling drugs menadione and paraquat (Storz *et al* 1987, Greenberg and Demple 1988). In protecting *E. coli* and *S. typhimurium* against oxidative damage, Ahp plays an analogous role to the seleno-cysteine glutathione peroxidases of eukaryotes, which also serve to protect cells from alkylhydroperoxides (Claiborne *et al* 1992) and have been found to be lacking in bacterial cells. It will therefore be interesting to know whether enzymes similar in structure and function to Ahp are widespread in bacterial species.

Figure 1.3

Schematic representation of the proposed mechanism of peroxide reduction by Ahp (after Jacobson *et al* 1989 and Claiborne *et al* 1992). For details see text.

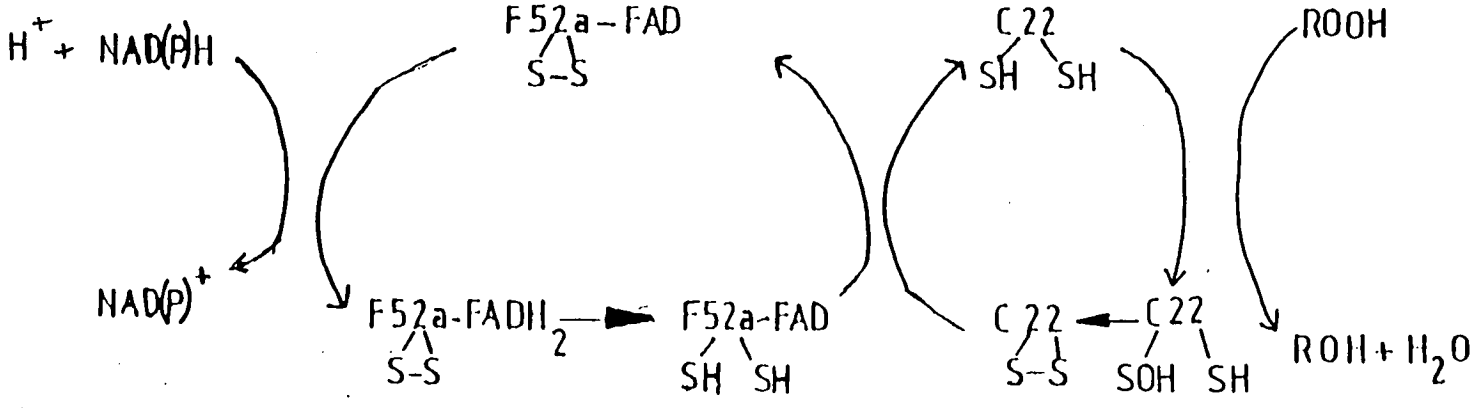

H+ + NAD(P)H
NAD(P)+
F52a-FAD
S-S
F52a-FADH2
S-S
F52a-FAD
SH SH
C22
SH SH
C22
S-S
C22
SOH SH
ROOH
ROH + H2O

FIGURE 1.3

1.10.6 HPI Catalase

Two distinct catalases have been characterised in *E. coli*. They are monofunctional HPII catalase (which is regulated in a growth phase dependent manner, by the *rpoS* encoded stationary phase sigma factor) and the bifunctional HPI catalase/peroxidase which is induced by peroxide stress in an OxyR-dependent manner.

HPI catalase is encoded by the *katG* locus which is located near 89.2 minutes on the standard *E. coli* linkage map (Loewen *et al* 1985) and encodes a 726 aa polypeptide of estimated MW 80049 (Triggs-Raine *et al* 1988). The enzyme exists as a homotetramer containing two molecules of protohaem IX, and rapidly dismutes H_2O_2 to O_2 and H_2O at a rate of 9.8×10^5 mol of H_2O_2 per mol of enzyme per minute (Claiborne and Fridovich 1978).

Comparative sequence analysis has shown HPI catalase to be distinct from the well characterised catalases of *S. cerevisiae, Penicillium* and humans, but structurally and functionally homologous to a group of bacterial catalase/peroxidase enzymes, for example sharing 48% homology with the thermostable peroxidase from *B. stearothermophilus* (Triggs-Raine *et al* 1988, Loprasert *et al* 1989). The slow peroxidatic activity displayed by HPI catalase requires exogenous electron donors and is unlikely to be of high significance *in*

vivo, as no substrate has been found in *E. coli* cell extracts which can outcompete the irreversible inhibition of the peroxide activity by diansidine. The catalase activity is also more advantageous to the cell as the disproportionation of H_2O_2 by this route requires neither exogenous electron donors nor energy input in the form of ATP, both of which are depleted during oxidative stress conditions. The peroxidase activity of HPI on the other hand requires electron donors, and would provide relatively poor protection within energy-depleted cells.

As with Ahp, extragenic suppression of the H_2O_2-hypersensitivity and high mutation rate of oxyR deletion strains of *E. coli* has been associated with mutations in the *katG* locus (Greenberg and Demple 1988). This emphasises the importance of HPI catalase in the elimination of H_2O_2 from *E. coli* cells both during normal aerobic growth and upon the onset of oxidative stress. Interestingly, the cellular location of HPI catalase is in the periplasmic space, whilst that of HPII catalase is cytoplasmic, suggesting that sources of H_2O_2 during exponential and stationary phase physiology may be different.

1.10.7 Glutathione Reductase

Glutathione reductase is a flavoprotein oxidoreductase

which catalyses the reduction of the oxidised form of glutathione (GSSG) by NADPH (Greer and Perham 1986). The enzyme helps to maintain intracellular levels of reduced glutathione (GSH). Moreover glutathione reductase activity is elevated in $OxyR^{(cons)}$ mutants of *S. typhimurium*, suggesting a role for GSH and glutathione reductase in oxidative stress (Christman *et al* 1985). GSH has been shown to reduce H_2O_2, O_2^- and HOO, generating a stable glutathione radical ($GS^{\cdot}$) which subsequently dimerises to form GSSG. It also promotes the reversal of disulphide bonds generated by protein oxidation (Farr and Kogoma 1991 and references therein).

Despite the likely benefits of maintaining adequate levels of GSH within oxidatively stressed cells, *gshA* mutants, which cannot synthesise glutathione, are no more sensitive to H_2O_2 than wild type cells, and are only slightly more sensitive to O_2^- generators (Greenberg and Demple 1986). The possible role of glutathione reductase during peroxide stress in *S. typhimurium* is therefore unclear. Perhaps its importance is masked by the compensatory action of other cellular antioxidants such as thioredoxin.

1.10.8 Other OxyR Regulated Genes

The *oxyS* gene of *E. coli* is divergently transcribed from and regulated by the product of the *oxyR* locus, near

89.6 minutes on the bacterial chromosome (Tartaglia *et al* 1992). It encodes a small 107 nucleotide non-coding RNA molecule which acts as a global activator or repressor of the expression of various genes during conditions of peroxide stress (personal communication from G. Storz). For example the *uhpT* gene, encoding a hexose-6-phosphate transporter, is induced by OxyS, possibly to meet the extra requirement for NADPH within oxidatively stressed cells, whilst *fhlA*, encoding a transcriptional activator involved in regulation of anaerobic metabolism, is repressed by OxyS. Over-expression of OxyS causes cell filamentation. The mechanism of OxyS regulation has not yet been elucidated but should provide further insight into the nature of the peroxide stress response.

Tartaglia *et al* (1992) have recently located an OxyR regulated promoter, several hundred base pairs upstream of the *ahp* operon of *S. typhimurium*. The promoter appears to direct transcription of an open reading frame (*orfO*) which is transcribed divergently from the *ahp* operon. The nature and function of the *orfO* gene product have not been reported.

Ribonucleotide diphosphate reductase may be partly regulated by OxyR (Farr and Kogoma 1991). The enzyme produces 2'-deoxyribonucleotide diphosphate precursors for DNA synthesis by catalysing the reduction of the

corresponding ribonucleotide diphosphates by thioredoxin and glutaredoxin. The stimulation of ribonucleotide diphosphate reductase production by OxyR may fulfil a requirement for extra DNA synthesis in response to peroxide-induced DNA damage.

Finally, proteases which specifically degrade oxidised proteins have been identified in *E. coli* (Davies *et al* 1988). Analysis of protease activity in an $OxyR^{(cons)}$ *E. coli* strain, revealed a 2 fold increase over the levels detected in $OxyR^{+}$ strains, suggesting that OxyR may regulate the production of the above proteases.

1.10.9 The OxyR Regulon and Membrane Repair

Farr *et al* (1988) have found evidence that *E. coli* cells have an inducible membrane repair capability, specific for H_2O_2-induced membrane damage. The authors measured the effects of sublethal concentrations of H_2O_2 or plumbagin on the functioning of the proton motive force-dependent and-independent membrane transport systems. Naive cells showed rapid loss of both functions upon treatment with H_2O_2 or plumbagin, requiring a recovery period of 90-100 minutes before full transport was restored. Cells which had been pre-adapted by exposure to low doses of H_2O_2 insufficient to impair membrane transport, recovered both proton motive force-dependent and-independent transport functions within 5-10 minutes

of a challenge with the damaging levels of H_2O_2. Subsequent experiments with the OxyR$^-$ and *katG* mutants demonstrated that the OxyR regulon was required for membrane repair, and that HPI catalase is essential but not sufficient for the restoration of membrane transport functions during peroxide stress. Ahp may also be involved in the membrane repair function, through the reduction of fatty acid hydroperoxides which are generated by lipid peroxidation (Farr and Kogoma 1991). Lipid peroxidation increases membrane fluidity, causing loss of structural integrity and function. The ability of Ahp to reduce lipid hydroperoxide is therefore likely to be important in the restoration of membrane transport.

The inducible membrane repair function described above is specific for H_2O_2 and cannot be induced by superoxide radical generators (Farr *et al* 1988). Similarly cells pre-adapted by H_2O_2 treatment are incapable of rapidly restoring plumbagin-generated loss of membrane transport, suggesting that oxidative membrane damage caused by H_2O_2 is distinct from that caused by superoxide radical generators.

1.11.1 Superoxide Stress and the Sox RS Regulon

As with H_2O_2, *E. coli* can mount an adaptive response to oxidative stress caused by superoxide radicals (O_2^-). This response was first demonstrated by Hassann and

Fridovich (1977) and Greenberg and Demple (1989), who found that cells pre-exposed to low doses of the superoxide radical generators PQ and MD became resistant to subsequent lethal challenges with the same agents. Pretreatment with superoxide radical generators also enhances the ability of cells to reactivate bacteriophage particles previously damaged *in vitro* by treatment with superoxide generators. This suggests that O_2^- induces a DNA repair pathway (Greenberg and Demple 1989). Pretreatment with superoxide radical generators does not enhance the resistance of *E. coli* to H_2O_2, nor does it induce a capability to reactivate H_2O_2-damaged bacteriophages (Greenberg and Demple 1989), indicating that the superoxide stress response is specific and distinct from that induced by H_2O_2.

2-D gel electrophoresis of [^{35}S]Met-labelled cell extracts reveals the complexity of the superoxide stress response. Greenberg and Demple (1989) found that PQ and MD induced almost all the peroxide stress proteins, including all the products of the OxyR regulon, plus an additional 33 proteins which were not inducible with H_2O_2. The authors suggested that the latter 33 proteins were directly induced by and therefore specific for O_2^- stress, whilst the peroxide stress proteins were induced by hydrogen peroxide produced through the action of superoxide dismutases (SODS). Walkup and Kogama (1989) avoided SOD-generated H_2O_2 by analysing the

superoxide stress response in *sodAB* double mutants, which lack both Mn SOD and Fe SOD activities. The lack of SOD activity allowed the *sodAB* cells to induce superoxide stress protein synthesis at ~1000 fold lower concentrations of superoxide generators than required in the wild type. The authors observed specific induction of ~30 proteins. Only one OxyR regulated protein - the F52a subunit of Ahp- was found to be induced by O_2^- in these *sodAB* mutants. Superoxide inducible proteins (SOIs) are listed in table 1.5.

A subset of SOIs is known to be under the positive control of the regulatory locus SoxRS (See Table 1.5). The SoxRS locus was identified by Greenberg *et al* (1990), who isolated *sox* constitutive ($sox^{[cons]}$), mutants which over-expressed ~nine SOIs, and Tsaneva and Weiss (1990) who isolated strains with extragenic mutations affecting the activity of an *nfo-lacZ* fusion. SoxRS is located near 92 minutes on the *E. coli* chromosome, close to the *ssb* gene which encodes the single stranded DNA binding protein. Tn*10* insertions in this region abolish the induction by superoxide of the SoxRS regulated proteins. Genes which are expressed constitutively in $Sox^{(cons)}$ mutants are *zwf* and *sodA* (output increased by 3-4 fold relative to Sox^+), *nfo* (5-8 fold increase), *soi*-17, *soi*-19 and *soi*-28 (Tsaneva and Weiss 1990). Modification of the ribosomal protein S6 is also elevated (Greenberg *et al* 1990).

Unlike OxyR$^{(cons)}$ mutants, Sox$^{(cons)}$ mutants generally do not show enhanced resistance to superoxide generating compounds, with the exception of *soxR101* and *soxR105* which were isolated as MD resistant mutants and are also resistant to PL but not PQ (Greenberg *et al* 1990).

TABLE 1.5 ***E. coli* GENES INDUCIBLE BY SUPEROXIDE STRESS**

Gene	Product	Sox RS Regulation	References
sodA	Manganese containing superoxide dismutase (MnSOD)	Yes	Carlioz & Touati (1986) Walkup & Kogoma (1989)
nfo	Endonuclease IV	Yes	Greenberg & Demple (1986) Walkup and Kogoma (1989)
zwf	Glucose-6-phosophate dehydrogenase	Yes	Greenberg & Demple (1986) Tsaneva & Weiss (1990)
ndh	NADH dehydrogenase	Not Known(a)	Reviewed by Farr & Kogoma (1991)
soi 28	Soi 28	Yes	Kogoma *et al* (1988), Walkup & Kogoma (1989)
soi 17/19	Soi 17/19	Yes	Kogoma *et al* (1988), Walkup & Kogoma (1989)
groES	GroES	Not known(a, b)	Greenberg & Demple (1986)
groEL	GroEL	Not known(a)	Vanbogelen *et al* (1987)
soxRS	SoxRS regulon activators	Yes	Tsaneva & Weiss (1990) Wu & Weiss (1992)
micF	Inhibitor of OmpF Translation	Yes	Greenberg *et al* (1990)
rimK	6 glutamic acid transferase	Yes	Greenberg *et al* (1990)

a) induced by Paraquat
b) induced by Menadione
General Reference: Farr and Kogoma 1991.

1.11.2 SoxRS Regulation

The finding that the SoxRS locus was very close to the *ssb* and *uvrA* genes on the *E. coli* chromosome enabled Tsaneva and Weiss (1990) to seek clones by virtue of this linkage, and then screen them for suppression of $soxR^{(cons)}$ mutations. Subsequent analysis of the proteins encoded by various subclones revealed that the *soxRS* region encoded two polypeptides of 13 and 18 kDa. Determination of the DNA sequence of the region revealed that the 17 and 13 kDa proteins were divergently transcribed from genes named *soxR* and *soxS* respectively (Amabile-Cuevas and Demple 1991, Wu and Weiss 1991). Comparative analysis of the predicted amino acid sequences of the SoxRS proteins revealed that SoxR is related to the MerR family of single component sensor-regulator proteins, sharing a common DNA binding domain and a conserved cluster of cysteines which may constitute the redox centre of a metal binding sensor domain. SoxS shares homology with the DNA binding segments of the AraC family of sensor regulator proteins, but is much smaller and lacks any homology with the sensor domain characteristic of other members of that family. Demple and Amabile-Cuevas (1991) proposed that SoxR and SoxS are the sensor and activator respectively of a two component regulatory system which governs the expression of the SoxRS regulon. Their model predicts that the SoxRS regulon is induced in two

stages: O_2^- activates SoxR (directly or indirectly) to stimulate SoxS transcription, which in turn activates transcription of the SoxRS regulon genes. The model is supported by the following evidence:-

a) Induction of the SoxRS regulon with paraquat stimulates the transcription of *soxS* as indicated by the 47-76 fold induction of a *soxS-lacZ* fusion (Wu and Weiss 1991).

b) Over-expression of SoxS from an inducible promoter actives expression of the SoxRS regulon in the absence of superoxide radical generators (Amabile Cuevas and Demple 1991).

c) A null mutation which inactivates SoxR but not SoxS results in loss of inducible paraquat resistance, and abolishes both paraquat induced *soxS* transcription and induction of the SoxRS regulon (Wu and Weiss 1992).

1.11.3 SoxQ Locus

Greenberg *et al* (1991) have isolated strains with mutations at a locus named SoxQ, near 34 minutes on the *E. coli* chromosome, which are phenotypically similar to *sox*$^{(cons)}$ mutants. *soxQ* mutations cause overexpression of Mn SOD, Glucose-6-phosphate dehydrogenase and *soi*-17, plus at least 7 other proteins not inducible by

oxidative stress. SoxQ+ is not required for PQ induction of the SoxRS regulon, but may provide an additional tier of control over this regulon.

1.11.4 SoxRS Regulation of *sodA*

E. coli contains two superoxide dismutases (SODs): a manganese containing SOD (MnSOD) encoded by *sodA* near 88 minutes on the standard linkage map, and the *sodB*-encoded iron containing SOD (Fe SOD). Only *sodA* is regulated by the *soxRS* system.

The function of SODs is to minimise the level of O_2^- within cells during both normal aerobiosis and oxidative stress. This is achieved by the rapid dismutation of O_2^- to H_2O_2, which is then eliminated by catalase activity. The importance of SOD activity was demonstrated by studies which monitored the steady state level of O_2^-, within *sodAB* double mutants, during normal aerobic growth. O_2^- was found to be present in *sodAB* mutants at concentrations ~10^{11} times that found in wild type cells.

Regulation of *sodA* expression is complex and is affected by multiple regulatory systems which couple the level of SodA to intra-and extra-cellular conditions which may lead to O_2^- production.

In addition to its activation by SoxRS and SoxQ, *sodA* is

repressed during anaerobiosis by the ArcAB two component regulatory system which controls the expression of respiratory enzymes under anaerobic conditions (discussed in Chapter 4 below). Repression of *sodA* by ArcAB is enhanced by the action of two more regulators of transcription - FNR a second major regulator of anaerobic gene expression, and integration host factor (IHF) - and serves to couple *sodA* expression to oxygen availability (Tardat and Touati 1993).

Expression of *sodA* is also coupled to intracellular iron levels by the Fur transcriptional regulator, which controls the expression of genes involved in iron uptake. Transient overloads in intracellular iron often occur during iron assimilation due to a delay in the shut off of iron uptake when normal iron concentrations have been restored. Such iron overload can result in oxidative stress due to reactions of iron with reactive oxygen species to produce OH and perferryl radicals (see Section 1.9.2 above). Therefore derepression of *sodA* simultaneously with iron uptake genes allows the potential oxidative threat of iron overload to be countered by a reduction in the availability of O_2^- (reviewed by Tardat and Touati 1993).

Regulation of *sodA* by SoxRS is independent of ArcAB, FNR and Fur as shown by the PQ inducibility of a *sodA-lacZ* gene fusion in strains carrying single, or combinations

of, mutations in the *fur, fnr* or *arcAB* genes (Hassan and Sun 1992).

1.11.5 Other SoxRS Regulated Functions

The elevation of glucose-6-phosphate-dehydrogenase expression by *soxRS* serves to enhance the production of NADPH during oxidative stress. NADPH is the electron source for the important antioxidant enzymes thioredoxin reductase, glutathione reductase and Ahp (although it can also utilise NADH *in vitro*). NADPH is also required by methionine sulphoxide reductase, which functions to repair oxidatively damaged proteins. Elevation of NADPH, rather than NADH, as a source of reducing equivalents avoids excess generation of hydroxyl radicals via the Fenton reaction, as NADPH unlike NADH does not reduce Fe^{3+} (reviewed by Farr and Kogoma 1991). Indeed an NADH diaphorase activity, which is elevated by O_2^- stress and has been linked with the *ndh* gene (encoding NADH dehydrogenase), serves to reduce the level of NADH within stressed cells (Farr and Kogoma 1991).

The *rimK* gene encodes a ribosomal modification protein which catalyses the addition of glutamyl residues to the ribosomal protein S6A, converting it to the S6C form. The observed rise in the level of S6C in *soxRS*$^{(cons)}$ mutants is therefore attributable to *rimK*

activity. Whilst this appears to mediate the resistance of *E. coli* to several antibiotics, it is not clear whether it has any role in oxidative defence.

The *micF* gene of *E. coli* is positively affected by SoxRS. It encodes an anti-sense RNA molecule which inhibits translation of *ompF* mRNA. OmpF is an outer membrane porin which facilitates the entry of small molecules into the cell (Nikiado and Vaara 1987). Reducing the level of OmpF results in increased resistance to a number of antibiotics and may inhibit the entry of naturally occurring redox active compounds, thus preventing O_2^- generation. Endonuclease IV (encoded by *nfo* which is inducible by SoxRS) aids repair of DNA strand breaks by removing 3' blocking groups to permit DNA polymerase I activity (Demple *et al* 1986). The role of Endo IV is therefore similar to that of exonuclease III during peroxide stress. However, their substrate specificities do not appear to be identical as Endo IV overproduction cannot fully suppress *xthA* mutations (Demple *et al* 1986).

1.12 Sigma -S and Oxidative Stress

Stationary phase *E. coli* cells express resistance to many causes of stress, including H_2O_2, which is mediated (at least partly) by the stationary phase sigma subunit encoded by *rpoS* (See Section 1.8.2). Two *rpoS* regulated genes have so far

been implicated in the resistance of stationary phase *E. coli* to H_2O_2: *katE* encoding HPII catalase and *xthA* encoding exonuclease III.

The *katE* gene has been mapped, cloned and sequenced (Loewen 1984, Von Ossowski 1991) and is located near 37.2 minutes on the *E. coli* chromosome. The native HPII catalase is a homohexamer containing 1 haem D group per 93000 Dalton subunit (Loewen *et al* 1986, Chiu *et al* 1989) and is monofunctional. Comparative sequence analysis revealed HPII catalase to be related to a group of catalases isolated from mammals, fungi and plant species and distinct from the HPI catalase encoded by *katG*.

Exonuclease III, encoded by *xthA* near 38 minutes on the genetic map, is an important DNA repair enzyme which removes 3′ blocks from DNA that has been nicked by H_2O_2 treatment (Demple *et al* 1986). *xthA* mutants are hypersensitive to H_2O_2 and NUV irradiation, indicating that blocking groups (sugar fragments) occur often at the 3′ termini of DNA strands broken by H_2O_2, thereby preventing DNA repair synthesis. Experiments conducted by Demple *et al* (1986) showed that *xthA* mutants exposed to H_2O_2 accumulated more DNA strand breaks than wild type cells and that DNA isolated from mutants was a poor substrate for DNA polymerase I repair activity, due to the presence of 3′ blocking groups at strand breaks. Prior incubation of H_2O_2 damaged DNA with exonuclease III facilitated DNA polymerase I mediated repair by removal of the 3′ blocking

groups.

Both HPII catalase and exonuclease III are regulated by *rpoS* in a growth phase dependent manner. HPII synthesis and activity have been shown to increase upon entry into early stationary phase (Mulvey *et al* 1990) whilst exonuclease III expression is positively affected by sigma S during late exponential phase (Sak *et al* 1989). Tn*10* insertions within *rpoS* abolish the stationary phase elevation of HPII catalase synthesis, reduce cellular levels of exonuclease III to those exhibited by *xthA* mutants, and sensitize cells to H_2O_2 and NUV irradiation. Another *rpoS* regulated gene which may be involved in protection against oxidative stress is *dps*, which encodes a histone-like DNA binding protein near 18 minutes on the *E. coli* chromosome (reviewed by Hengge-Aronis 1993).

Dps, which is a major protein late in stationary phase, binds tightly to DNA *in vitro* and may perhaps help prevent oxidative DNA damage. A second function of Dps - DNA interaction may be global regulation of gene expression. Recently, Storz and co-workers found that *dps* transcription may be positively regulated by OxyR (personal communication from G. Storz). This may represent the first example of overlapping regulation between the peroxide stress and stationary phase responses. To my knowledge no other evidence exists of regulatory overlap between stationary phase gene expression and other oxidative defence regulons.

Chapter Two

Materials and Methods

2.1 MATERIALS

2.1.1 Chemicals and Enzymes

Most chemicals were of analytical reagent 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.

Fison plc, Loughborough, Leicestershire.

Sigma Chemical Company, Poole, Dorset.

Agarose for gel electrophoresis of DNA was obtained from Miles Scientific. Low electroendosmosis agarose 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 and radiolabelled compounds were obtained from Amersham International plc, Amersham, Buckinghamshire.

Enzymes used for nucleic acid manipulation were purchased from:-

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

New England Biolabs, Beverly, MA, USA.

Pharmacia LKB Biotechnology, Milton Keynes, Buckinghamshire.

United States Biochemical Corporation, Cleveland, Ohio, USA.

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

Rabbit anti-sigma peptide antiserum was a kind gift from Professor Akira Ishihama, National Institute of Genetics, Mishima, Japan.

2.1.2 Standard Solutions

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

Solutions were stored at room temperature, unless stated otherwise.

1 x TE	:	10 mM Tris-HCl, 1mM EDTA; adjusted to pH8.0 with concentrated HCl.
10 x TAE	:	0.4M Tris base, 0.2M sodium acetate, 10mM EDTA; adjusted to pH8.3 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; pH8.2 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 pH7.0.
20 x SSPE	:	3.6M NaCl, 0.2M NaH_2PO_4, 20 mM EDTA; adjusted to pH7.4 with 5M NaOH.
50 x Denhardt's Solution	:	1% (w/v) Ficoll, 1% (w/v) polyvinylpyrrolidine, 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 pH8.6.
RNase A	:	5mg ml^{-1}. Heated to 100°C for 15 minutes, then cooled to room temperature before storage, in aliquots, at -20°C.
Phenol	:	Redistilled under nitrogen into TE. Equilibrated with 1M Tris-HCl, then TE buffer. 8-hydroxyquinoline was added to 0.1% (w/v) final volume and

aliquots stored at -20°C. Phenol equilibrated with dH_2O was used for extractions of RNA solutions.

Chloroform : Chloroform and isoamylalcohol mixed 24:1 (v/v); stored in darkened bottles at room temperature.

IPTG : 25mg l^{-1} stored at -20°C.

X-gal : 25mg ml^{-1} in dimethyl formamide. Stored at -20°C.

DMSO : A freshly opened bottle of spectroscopic grade dimethyl sulphoxide was dispensed under nitrogen into 20ml plastic universal bottles. These were stored at -70°C and dispensed into 1ml aliquots in microfuge tubes, as required.

2.1.3 Growth Media

Growth media were made up with dH_2O and sterilized by autoclaving or filter sterilized. 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 pH7.2 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.

Dye Agar : 10g Difco Bactotryptone, 8g NaCl, 200mg toluidine blue or methylene blue.

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, 20ug ml^{-1} amino acids, 2ugml^{-1} 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 Dye agar and minimal media and agars were prepared by the staff of the Media Room of this Institute.

2.1.4 Antibiotic Stock Solutions

Ampicillin : 10mg ml^{-1} in dH_2O.

Chloramphenicol : Free chloramphenicol (20mg ml^{-1}) in ethanol.

Kanamycin : 20mg ml^{-1} kanamycin sulphate in dH_2O

Streptomycin : 80mg ml^{-1} streptomycin sulphate in dH_2O.

Tetracycline : 20mg ml^{-1} in 50% ethanol.

Antibiotic solutions were filter sterilized and stored in aliquots at -20^0C. The required amount of stock solution was added to liquid media at room temperature and to molten agar media (cooled to 55^0C) immediately prior to pouring.

2.1.5 Bacterial Strains

All strains were derivatives of *E.coli* K12. with the exception of the *S. typhimurium* strain LT2, and are described 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 recA13/F' proAB*$^{+}$ *lacI*^{q} *lacZ*ΔM15	N. E. Murray (Edinburgh)
TG1	Δ(*lac-pro*) *thi supE hsdDS/F' traD36 proAB*$^{+}$ *lacI*^{q} *lacZ*ΔM15	Gibson (1984)
NM621	*recD hsdR mcrB pro leu thi lac tonA21 supE*	N. E. Murray
DSA100	NM621 *scrP::kan*	This work
MM38	*argG6 hisG1 leuB6 metB1 pyrE gal-6 lacY1 xyl-7 supE44 bgl*$^{+}$ *tonA2 gyrA polA*(Am) *rpsL104 tsx-1 uhp*	M. Masters (Edinburgh)
DSA101	MM38 *scrP::kan*	This work
DSA102	NM621 Δ(*ahpC-F'*) *kan::ahpF*	This work
A19	HfrPO21 *metB1* (λ) *rel-1 his*	Kindler *et al* (1973)
DSA103	A19 Δ(*ahpC-F'*) *kan::ahpF*	This work
P678.54	*thr leu thi minB*	K. Begg (Edinburgh)
DS410	*minB thi ara gal xyl mtl tonA rpsL*	Dougan & Sherratt (1977)
S159	*uvrA thi*	Jaskunas *et al* 1975
S159 (lambda *ind*)	S159 Lysogenised with lambda *ind*	Jaskunas *et al* 1975
ECL585	Δ*frd101 araD139* Δ(*argF-lac*) *rpsL150 relA1 deoC1 flb-5301 ptsF25* Φ(*sdh-lac*) *arcA1 zjj::*Tn*10*	Iuchi & Lin (1988)
W3110	*SupE?*	Bachman (1972)
LT2	wild type *S. typhimurium*	M. Gallacher (Edinburgh)

TABLE 2.2 BACTERIOPHAGES

Bacteriophage	Description	Reference
Lambda 166	W3110 genomic DNA from 643 Kb - 661 kb[a] cloned between the *Bam*HI sites of lambda EMBL4	Kohara *et al* (1987)
Lambda 167	W3110 genomic DNA from 651 Kb - 668 kb[a] cloned between the *Bam*HI site of lambda EMBL4	Kohara *et al* (1987)
Lambda 522	W3110 genomic DNA from 3404 Kb - 3420 kb[a] cloned between the *Bam*HI site of lambda EMBL4	Kohara *et al* (1987)
Lambda 523	W3110 genomic DNA from 3410 Kb - 3429 kb[a] cloned between the *Bam*HI sites of lambda EMBL4	Kohara *et al* (1987)
M13 mp18	*lac Z*, cloning vector	Yanisch-Perron *et al* (1985)
M13 mp19	As mp18 but with the cloning sites in the opposite orientation	Yanisch-Perron *et al* (1985)
M13 mp18-15	4 kb *Eco*RI-*Hind*III fragment from lambda 522 cloned between the same sites in M13 mp18	Fiona Townsley (This work)
M13 mp19-23	4 kb *Eco*RI-*Hind*III fragmented from M13 mp18-15 cloned between the same sites in M13 mp19	This work
M13 mp19-22	4.1 kb *Hind*III-*Kpn*I fragment from lambda 166 cloned between the same sites in M13 mp19	This work

Inconsistent in kb - J.F prefers Kb

TABLE 2.2 BACTERIOPHAGES (Continued)

M13 mp19-19	1.04 kb *Eco*RI fragment from M13 mp19-22 cloned into M13 mp19 cloned into M13 mp19	This work
M13 mp19-20	1.8 kb *Pst*I fragment from mp19-22 cloned into M13 mp19	This work
M13 mp19-21	As M13 mp19-20 but with the *Pst*I fragment inserted in the opposite orientation	This work
M13 mp18-16	*Hin*dIII-*Sma*I fragment from M13 mp19-22 cloned in the same sites in M13 mp18	This work
P1vir	Generalised transducing phage used in strain constructions	Laboratory stock

a. numbers in kilobase pairs are co-ordinates on the physical map of the *E. coli* K12 genome (Kohara *et al* 1987)

TABLE 2.3 PLASMIDS

Plasmid	Description	Reference or Source
pACYC184	Cm^r, Tet^r, rep p15A cloning vector	Chang and Cohen (1978)
pSKS106	Amp^r, *lacZ*$^+$*Y*$^+$*A*$^+$ cloning vector	Shapira *et al* (1983)
pBluescriptIISK+	Amp^r, *lacZ* cloning vector	Stratagene Ltd. Cambridge UK
pJW11	Amp^r, *galK*$^+$ cloning vector	Wright (1987)
pUC4-KISS	Amp^r, Kan^r, source of Tn*901* kanamycin resistance cassette	Barany (1985)
pDSA23	4.17kb *Hin*dIII-*Bgl*II fragment from mp19-22 cloned between the *Hin*dIII and *Bam*HI sites of pACYC 184. Cm^r Tet^s *ahpC*$^+$*F*$^+$	This work
pDSA24	2.3kb *Sma*I fragment deleted from pDSA23, Cm^r *ahpC*$^+$	This work
pDSA25	2.3 kb *Sma*I fragment from pDSA23 cloned into pBluescriptIISK+ Amp^r, *ahpF*$^+$	This work
pDSA26	2.3 kb *Bam*HI-*Hin*dIII fragment from pDSA25 cloned between the same sites in pSKS106. Amp^r, *ahpF*$^+$	This work
pDSA27	1.3 kb *kan* cassette from pUC4-KISS cloned between *Pst*I sites in pDSA23, Amp^s, Kan^r Plasmid used for disruption of the *ahpCF* genes	This work
pFMT1	4 kb *Eco*RI-*Hin*dIII fragment from M13 mp18-15 cloned between the same sites in pJW11, Amp^r, *galK*$^+$, *scrP*	Fiona Townsley (This work)
pDSA28	1.3 kb *kan* cassette from pUC4-KISS cloned into the unique *Nsi*I site of pFMT1. Amp^r, Kan^r Plasmid for disruption of the *scrP* gene.	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 Pipetman P-20, P-200 and P-1000 air displacement pipettors and tips, or glass pipettes.

Microfuge tubes, pipette tips and glass pipettes used in procedures were sterilized by autoclaving. Corex tubes were sterilized by baking at 250°C for 16 hours. Falcon tubes were supplied presterilized 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 in dH_2O, then boiled 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 dH_2O 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 (pH5.2) 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^{-1} for dsDNA and 40µg ml^{-1} 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 (Du Pont Cronex or Amersham Hyperfilm MP) was exposed to gels or 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 using preflashed x-ray film.

[^{35}S] labelled nucleic acids and proteins in fixed and dried gels were autoradiographed at room temperature. For increased sensitivity, gels containing [^{35}S] labelled proteins were soaked in Enlightning (Du Pont) prior to drying, then autoradiographed at -70°C with preflashed film and intensifying screens.

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 appropriate antibiotics for maintenance of plasmids 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 relevant 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 indefinitely.

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 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 0.1M $CaCl_2$ (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.

Transfection and transformation were effected by adding DNA (50-200ng plasmid or M13 RF DNA, or ½ 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) 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 aliquots 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, then plated as for M13 titration (see Section 2.4.9).

2.3.3 Dye Sensitivity Tests (Iuchi and Lin 1990)

5ml overnight cultures of test and control strains were diluted in bacterial buffer to give 200-500 viable cells per 0.1ml. Diluted suspension (0.1ml per plate) was then spread on dye agar. After incubation at 37°C overnight dye sensitive strains gave rise to only very small colonies or no visible colonies. Dye resistant strains produced clearly visible colonies.

2.3.4 Peroxide Sensitivity Assays (Storz *et al* 1989)

0.1ml aliquots of fresh overnight cultures of the test

strains were mixed with 2.5ml aliquots of molten water top agar and poured onto LB agar, or Spizizen minimal agar plates. 10µl aliquots of cumene hydroperoxide (5% v/v in DMSO) were spotted onto 5mm paper discs, and the discs placed in the centre of the agar. After incubation at 37°C for 24 hours, the diameter of the zone of killing on each plate was recorded.

2.3.5 Test for RNase I (Gesteland 1966)

Single colonies were replica patched onto a pair of LB agar plates, and incubated overnight at 37°C. The next day one plate was stored at 4°C whilst the other was overlayed with 2.5ml of RNA top agar (0.7% {w/v} Difco agar, 3% {w/v} yeast RNA, 0.1M EDTA; pH7.0) and placed at 42°C. After 2-4 hours incubation, 3ml of HCL (0.1M) was poured onto the plate to precipitate the RNA. $RNaseI^+$ colonies were surrounded by clear "halos", within the white RNA precipitate, due to the degradation of RNA by RNaseI released into the top agar by EDTA treatment.

plate. The plate was incubated upright at 37°C overnight. To harvest the phages, the top agar was scraped off the plate using a sterile spatula and transferred to a 15ml Corex tube. 1.5ml of phage buffer and 0.15ml of $CHCl_3$ were added and the agar mashed by pipetting up and down several times using a 10ml pipette. The tube was then incubated with shaking at 37°C for 60 minutes. Thee phage suspension was clarified by centrifugation (16000g) in a Sorvall SS34 rotor at 4°C for 10 minutes and the cleared supernatant transferred to a fresh tube. A few drops of $CHCl_3$ were then added and the lysate stored at 4°C.

2.4.3 <u>Transduction with P1 (Modified from Miller {1972})</u>

Cells were grown and subcultured as for titrations. When an A_{650} of 0.5 was reached, cells were pelleted by centrifugation (2500g) for 5 minutes and resuspended in 1ml of the supernatant. 0.1ml aliquots of cells were then mixed with 5 x 10^6 p.f.u of P1vir grown on the appropriate donor and incubated at 37°C for 20 minutes. 0.2ml of trisodium citrate dihydrate (1 M) and 1ml of LB were then added and the mixture incubated at 37°C for two more hours to allow expression of antibiotic resistance (trisodium citrate chelates Ca^{++} ions, thus preventing reinfection of transductants with virulent P1). The mixture was then added to 2.5ml of Difco top agar and plated on selective medium. Candidate transductants

2.4 PHAGE TECHNIQUES

2.4.1 P1 Titrations

10ml of LB, supplemented with glucose (1% w/v) and $CaCl_2$ (5mM), were inoculated with 1/20 volume of an overnight culture of the host strain and incubated at 37°C with shaking. When the culture had reached an A_{650} of 0.5, cells were pelleted (2500g) in a MSE Minor 'S' bench-top centrifuge for 5 minutes, then resuspended in 5ml of sterile $MgSO_4$ (10mM), 5mM $CaCl_2$ and shaken at 37°C for 15 minutes.

0.2ml aliquots of these plating cells were mixed with 0.1ml aliquots of serially diluted P1 lysates and incubated for 15 minutes at 37°C to allow adsorption of the phage to bacterial cells. 3ml of BBL top agar containing $CaCl_2$ (5mM) were added and the mixture poured onto fresh, well dried BBL agar plates containing $CaCl_2$ (5mM). Plates were then incubated inverted at 37°C overnight.

2.4.2 P1 Plate Lysates

Host cells were prepared as for P1 titrations. 0.2ml of plating cells and 10^6 p.f.u of phage (in 0.1ml of phage buffer) were mixed and incubated at 37°C for 15 minutes. 3ml of BBL top agar containing $CaCl_2$ (5mM) were added and the mixture poured onto a fresh, moist LB

were then purified and characterised.

2.4.4 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 Minor 'S' 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 phage 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.

2.4.5 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 occurred (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.

2.4.6 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^{-1} 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.

2.4.7 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.

2.4.8 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 2.4.5, 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 a MOI of 1 and incubated with shaking at 37°C for 15 minutes. Trisodium citrate and glucose were then added to final concentrations 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.

2.4.9 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^{-1}) and X-gal (250µg ml^{-1}) 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.

2.4.10 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 2.4.9.

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 & 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 pH8.0, 10mM EDTA pH8.0.

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 lysozyme (4mg ml^{-1}) 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:isoamyl alcohol. Nucleic acids were concentrated by ethanol precipitation and vacuum dried.

The dried pellet was then redissolved in 30µl of TE containing RNase A (20µg ml^{-1}).

2.5.2 Large Scale Isolation of Plasmid DNA

200ml of LB, plus appropriate antibiotic, 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^{-1}) was added and the sample incubated for 5 minutes on ice. 1µl of 0.5M EDTA (pH 8.0) and 800µl of RNase A (5mg ml^{-1}) were added and the sample 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^{-1}) were added. The sample was then transferred to a crimp-seal polyallomer tube and centrifuged (90000g) 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 with 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 remove any remaining bacterial cells. The supernatant was transferred to a fresh 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

temperature. 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 2.4.7, 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 and once with chloroform:isoamyl alcohol. The aqueous phases were then pooled and dialysed against several changes of TE buffer at 4^{o}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 (Modified from Hermann and Frishcauf 1987)

100ml of LB (plus appropriate antibiotic) were inoculated with 1/50 volume of an overnight culture and shaken at 37°C until A_{650} 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 20ml of STE (25% w/v Sucrose 50mM Tris-HCl pH8.0. 10mm EDTA pH8.0) and transferred to a Falcon tube. 1ml of SDS (10% w/v) and 0.4ml of proteinase K (10mg ml^{-1}) 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 inverting gently 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 dissolution was complete.

2.5.7 Purification of RNA from *E. coli* Cells (Salser *et al* 1967, 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. Where appropriate, the culture was treated with H_2O_2 (at 65µM final concentration) and grown for a further 10 minutes. 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 (pH7.3) and transferred to a microfuge tube. Lysozyme was then added to a final concentration of 300µg ml^{-1} 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 (pH5.2) 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 pH7.5, 10mM NaCl.

2.6 MANIPULATION OF DNA

Unless otherwise stated, methods described in the following sections are modifications of methods described by 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 generally performed in 1 x Universal buffer (33mM Tris-acetate pH7.9, 10mM Mg-acetate, 66mM K-acetate, 5mM DTT, 0.1mg ml^{-1} 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 of 0.6-2% (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 (50 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 14cm) 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 2.1.3).

After electrophoresis, gels were stained in EtdBr (1μg ml^{-1}) 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, using Polaroid 667 positive film.

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, California, USA).

Briefly: The DNA was digested with the appropriate enzyme(s) and the products separated by electrophoresis in a TAE agarose gel (see Section 2.6.3). The EtdBr stained gel was examined on a UV transilluminator 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 5ul) 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 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 (pH7.6), 0.1M Mg Cl_2, 0.1M DTT, 500 µg ml^{-1} BSA, 10mM ATP), DNA ligase and dH_2O.

For ligation of cohesive ends, 0.5 - 2 units of T_4 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 2 - 5 units of T_4 DNA 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 radio-labelled. This produces a population of radio-labelled DNA molecules ranging from a few nucleotides to hundreds of nucleotides in length. In the second step, four parallel reactions are performed, one for each of the four bases (A, C, G and T). DNA synthesis is continued in each reaction, in the presence of higher concentrations of all four dNTPs plus one 2:3'-dideoxyNTP. 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 pH7.5, 100mM $MgCl_2$, 250 mM 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 [^{35}S] dATP alpha S (10µCi ul^{-1}) and 2µl of labelling mix (1.5uM each of dGTP dCTP and dTTP). 2µl of Sequenase T7 DNA polymerase (diluted 1/8 in 10mM Tris-HCl pH7.5, 5mM DTT, 0.5 µg ml^{-1} BSA) were added and the mixture incubated for five minutes at room temperature.

Termination Reactions

Four microfuge tubes were labelled A, C, G and T. 2.5µl of the appropriate termination mix (Table 2.5) were added to each tube and the tubes transferred to a 37°C water bath 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 five minutes. The reactions were stopped by addition of 5ul of formamide dye (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cycanol FF). The tubes

were then transferred to a 75°C water bath for 2-5 minutes and loaded immediately onto a denaturing polyacrylamide gel (see Section 2.7.4).

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 0.2M NaOH, 0.2mM EDTA (10µl final volume) for 15 minutes at 37°C. The solution was neutralised by addition of 0.1 volumes of 3M sodium acetate (pH5.2) and the DNA precipitated with 2-3 volumes of ethanol. The DNA pellet was vacuum-dried, then resuspended in 10µl of 1 x Sequenase annealing buffer.

Annealing of primers to double-stranded templates was carried out at 37°C for 30 minutes. Sequencing was then performed as for single stranded templates.

2.7.3 Sequencing Using [^{32}P] End-Labelled Primers

DNA sequence was also determined using oligonucleotides labelled at their 5' ends with [^{32}P] ATP (as described in Section 2.8.2).

0.5pmol of labelled primer and 1-3µg of template DNA

were annealed in the usual way. 2.5µl of dH_2O, 1µl of DTT(0.1M) and diluted Sequenase were added and the mixture split between four tubes containing pre-warmed termination mixes.

Termination reactions were allowed to proceed for 5 minutes at 37°C, then stopped by addition of formamide dye (5µl).

TABLE 2.4 TERMINATION MIXES FOR SEQUENCING

Nucleotide	A Mix	C Mix	G Mix	T Mix
dATP	80	80	80	80
dCTP	80	80	80	80
dGTP	80	80	80	80
dTTP	80	80	80	80
ddATP	8	-	-	-
ddCTP	-	8	-	-
ddGTP	-	-	8	-
ddTTP	-	-	-	8

All units shown are in μmol L^{-1}

2.7.4 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.

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 2 glass plates, one notched, one rectangular, 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. Where desired, sodium acetate (pH5.2) was added to the bottom tank to a final concentration of 0.3M. Adding sodium acetate results in slower migration of DNA molecules in the bottom of the gel. This lessens the vertical spacing between adjacent bands, allowing the sequence near the bottom of the gel to be read more easily (personal communication from B. Grimes). 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 3 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.

Gels containing [^{32}P] nucleic acids were fixed and dried

when fine resolution was required, otherwise they were wrapped in Saran Wrap and autoradiographed directly without fixing or drying.

2.8 NUCLEIC ACID HYBRIDISATION

2.8.1 Labelling of DNA Fragments as Probes by Random Priming

Isolated DNA fragments were labelled to high specific activity by the method of Feinberg and Vogelstein (1984). The following solutions were used in this method:

Solution A: 1.25M Tris-HCl (pH8.0), 0.125M $MgCl_2$, 0.5mM each of dATP, dGTP, dTTP, 0.25mM 2-mercaptoethanol.

Solution B: 2M Hepes-NaOH (pH6.6).

Solution C: 4.5mg ml^{-1} random hexadeoxyribonucleotide mixture (in TE).

Digested DNA was electrophoretically separated in an 0.6% low melting point (LMP) agarose gel (Bethesda Research Laboratories). A gel slice containing the desired fragment was excised from the EtdBr stained gel and weighed. dH_2O was added (3ml per gram of gel) and the sample boiled to melt the agarose and denature the DNA. The solution was either used immediately or stored at -20°C. DNA prepared in this way can be labelled

directly without further purification.

To label the DNA the following reagents were mixed in a microfuge tube:

dH_2O to a final volume of 50ul, 10ul of oligonucleotide labelling buffer (solutions A:B:C mixed 2:5:3), 1µl of BSA (20mg ml^{-1}) 20-100ng of DNA in LMP agarose, 50 µCi of alpha [^{32}p]-dCTP and 2 units of *E. coli* DNA polymerase Klenow fragment. The reaction was allowed to proceed at room temperature for 2-18 hours, then terminated by addition of "stop" solution (20mM NaCl, 20mM Tris-HCl pH7.5, 2mM EDTA, 0.25% (w/v) SDS, 1µM dCTP). The labelled probe was denatured before use by boiling for 3 minutes.

2.8.2 Labelling of Oligonucleotide Probes

15pmol 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), 2.5µl [gamma^{32}P]ATP (10mCi ml^{-1}, 3000 Ci $mmol^{-1}$), 20 units T4 polynucleotide kinase and dH_2O to 25ul. The mixture was incubated at 37°C for 30 minutes then the enzyme inactivated by heating at 75°C for 15 minutes. The labelled oligonucleotide was then stored at -20°C until needed.

2.8.3 Transfer of DNA to Nylon Membranes (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 in the usual way, 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 pH7.2, 1mM EDTA) for 15 minutes each, and finally rinsed in dH_2O.

To prepare the blotting platform a glass plate, supported by 2 inverted Petri dishes, 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 3 sheets of Whatman 3MM paper soaked in blotting buffer and a stack of absorbent paper towels. The gel was then surrounded by Cling-Film 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 (0.5kg) weight. Transfer of DNA from the gel to the nylon membrane was allowed to proceed for 2-4 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 Cling-Film and stored at room temperature.

2.8.4 Hybridization of DNA Probes to DNA Immobilised on Nylon Membranes

Membranes prepared as in Section 2.8.3 were prehybridised with 20-25ml of SSPE hybridisation buffer (6 x SSPE, 50% v/v formamide, 5 x Denhardt's solution, 0.5% w/v SDS, 0.5mg ml^{-1} heparin from porcine intestinal mucosa) in heat sealed plastic bags at 65°C for two hours. 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 due to sequence-specific binding of the probe.

After prehybridisation the bag containing the membrane was cut open at one corner and the hybridisation solution discarded. Fresh hybridisation buffer

containing the radiolabelled probe was added and the bag resealed. Hybridisation was allowed to proceed, with shaking, at 65°C for 12-18 hours. The hybridisation solution was then removed from the bag and stored at -20°C for re-use. The membrane was then washed briefly at room temperature in 6 x SSPE, then twice at 65°C, for 20 minutes each, in 1 x SSPE, 0.2% w/v SDS. The membrane was then wrapped in Saran Wrap and autoradiographed for 12-24 hours in the first instance, then longer where necessary.

2.8.5 Hybridisation of Oligonucleotide Probes to DNA Immobilised on Nylon Membranes

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

$$Tm\ (°C) = 4\ (G + C) + 2\ (A + T)$$

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

(A + T) = number of A : T base pairs.

Filters were prehybridised with 20-25ml of SSC hybridisation buffer (6 X SSC, 0.5% w/v SDS, 50mM sodium pyrophosphate, 0.5mg ml^{-1} heparin from porcine intestinal mucosa) for 2 hours at 10°C below the estimated Tm. 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, 0.2% w/v SDS, for 15 minutes each, at 30°C. Finally, the filter was washed twice in 6 X SSC, 0.2% w/v SDS at Tm - 10°C for five minutes each, then wrapped in Cling-Film and autoradiographed.

Where more stringent conditions were required, filters were washed twice for five minutes each in 2 X SSC, 0.2% SDS at Tm - 10°C.

2.8.6 Removal of Probes from Nylon Membranes

DNA probes and oligonucleotide probes can be removed from a nylon membranes 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 completely. If clean it was then allowed to dry at room temperature, then wrapped in Saran Wrap and stored at room temperature. Membranes treated in this way were prehybridised and hybridised with new probes in the normal way.

2.8.7 Primer Extension Analysis of RNA (Modification of the Method described by Sambrook *et al* 1989)

Primer extension analysis 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.

20-60ug RNA prepared as described in Section 2.5.7 were mixed with 0.3-1.2 pmol oligonucleotide primer (end-labelled as described in Section 2.8.2). 3µl formamide (special biochemical grade, BDH), 6ul of 5 x reverse transcription buffer (250mM Tris-HCl pH7.5, 375mM KCl, 15mM $MgCl_2$) and dH_2O to 30ul 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 Tm 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 reverse transcription buffer, 5µl DTT (0.1M), 1µl of dNTP mix (25mM each of dATP, dGTP, dCTP, dTTP), 1µl of BSA (5mg ml^{-1}), 100 units of M-MLV RTase (Mouse Moloney Leukemia Virus Reverse Transcriptase, Gibco BRL) and dH_2O to 50ul final volume. The mixture was then incubated at 42°C for 60 minutes.

After this time, 12.5µl NaOH (0.5M) were added and the sample boiled for three minutes to hydrolyse RNA. 12.5µl each of HCl (0.5M) and 1M Tris-HCl (pH7.4) were added and the sample precipitated with ethanol. The dried DNA pellet was dissolved in 4µl dH_2O and mixed with 6ul formamide dye (see Section 2.7.1). Samples were then boiled for three minutes and the cDNA products size-separated on DNA sequencing gels (see Section 2.7.4). A DNA sequence ladder, generated using the same end-labelled primer, was run alongside as a marker.

2.9 EXPRESSION AND LABELLING OF PLASMID-ENCODED PROTEINS IN MINICELLS (STOKER *et al* 1984)

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 into minicells quite efficiently; therefore purified minicells can be used to analyse plasmid-encoded protein synthesis under essentially *in vivo* conditions.

2.9.1 Purification of Minicells

Three 500ml batches of LB plus appropriate antibiotic(s) were inoculated with 1ml aliquots of a fresh overnight culture of the minicell-producing strains DS410 or P678.54, transformed with the required plasmid, and shaken overnight at 37°C. At the same time eight

sucrose gradients were prepared by freezing 35ml aliquots of 20% w/v sucrose in M9 minimal medium in 50ml polycarbonate centrifuge 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 (700g) in a Sorvall GS3 rotor for 15 minutes at 4°C. Minicells were then harvested from the supernatant by centrifugation (11000g) 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, in order to disperse aggregates of minicells which would otherwise pellet with remaining viable cells during 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 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) as calculated by the following equation:

$$V = \frac{0.2 \times 100}{A_{600}}$$

where V = Volume of the aliquot.

Aliquots were stored at -70°C.

2.9.2 Labelling of Plasmid-Encoded Proteins in Minicells

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. 20 µCi of L-[^{35}S] methionine (10 µCi ul^{-1}, 1000 Ci $mMol^{-1}$) in Difco methionine assay medium (25% w/v in M9 glucose) were added and the samples incubated at 37°C for three hours. Samples were then "chased" with 5ul unlabelled methionine (8mg ml^{-1}) for three minutes, and the minicells were pelleted by centrifugation for three minutes. Minicells were then washed in 50mM Tris-HCl

(pH 6.8), pelleted as before, and resuspended in residual liquid. 15µl of SDS sample buffer (50mM Tris-HCl pH6.8, 1% (w/v) SDS, 10% w/v glycerol, 0.8% w/v bromophenol blue, 1% 2-mercapto-ethanol) were added and the samples boiled for three minutes. Samples were then fractionated in SDS polyacrylamide gels as described in Section 2.9.10. Labelled proteins were detected by autoradiography.

2.9.3 Expression and Labelling of Proteins in UV-irradiated Host Cells (Jaskunas *et al* 1976, Stoker *et al* 1984)

When *E. coli* cells are heavily UV irradiated chromosomal DNA is damaged to such an extent that transcription of all but the smallest chromosomal genes is reduced to a negligible level. If cells treated in this way are then infected with phage lambda, phage-encoded proteins can be specifically labelled by incorporation of radiolabelled amino acids such as [^{35}S] methionine.

If the irradiated host is lysogenic for phage lambda, then transcription from the strong phage promoters, P_L and P_R, is repressed by the *c*I gene product. Only cloned genes served by strong bacterial promoters will be expressed under these conditions. (Phage encoded *c*I and *rex* gene products will also be expressed.) Where cloned genes are served by weak promoters or autogenous promoters are not present, non-lysogenic hosts can be used. In this case transcription is initiated at P_L

and P_R resulting in expression of both phage and cloned genes.

30ml M9 Maltose, supplemented with 10mM $MgSO_4$, were innoculated with *E. coli* strains S159 or S159 (lambda *ind*) to give an A_{600} of 0.1. The culture was grown to A_{600} of 0.6, then 15ml transferred to a polycarbonate centrifuge tube and spun (8000g) for 15 minutes at 15°C. The pellet was resuspended in 4.5ml ice cold 1 x M9 salts containing $MgSO_4$ (20mM) and methionine (0.01µg ml^{-1}). 4ml of suspension were transferred to a glass Petri dish on ice and UV irradiated for 7.5 minutes at 600 ergs/mm^2/min. The irradiated culture was then transferred to a small (25ml) foil-wrapped conical flask and shaken at 37°C for five minutes to allow dissipation of free radicals and decay of preformed mRNA. The foil excludes light, to prevent photo-repair of DNA damage.

50µl (5 x 10^7 cells) were then transferred to a microfuge tube containing 5 x 10^8 p.f.u. of the appropriate phage in 5µl phage buffer. The culture was then incubated for a further 10 minutes at 37°C to allow adsorption of phage particles to the cells. 0.2ml pre-warmed label mix (1 x M9 salts, 1mM $MgSO_4$, 0.01µg ml^{-1} carrier L-methionine and 100µCi ml^{-1} L-[^{35}S] -methionine) was then added and the culture incubated for three minutes at 37°C, then chased for three minutes with 30µl unlabelled methionine

(2mg ml^{-1}). The sample was then mixed with 10ul sodium azide (1M) and centrifuged for six minutes at 4°C. The pellet was resuspended in 0.5ml of 50mM Tris-HCl (pH 7.5), 200ug methionine ml^{-1}, 0.1mM DTT and centrifuged as before. The supernatant was discarded and the cells resuspended in residual liquid. 60µl of SDS sample buffer were added and the mixture boiled for three minutes. Samples were then fractionated on SDS polyacrylamide gels (Section 2.10) and the dried gels were autoradiographed.

2.9.4 Immunoprecipitation of Labelled Proteins

[^{35}S] Met-labelled proteins were selectively immunoprecipitated from *E. coli* crude cell extracts using rabbit anti-22-peptide antiserum (Fujita *et al* 1987) and Protein A-sepharose beads (Pharmacia LKB) as immunoadsorbant. The following solutions were used in this method:

Wash buffer I (WBI): 10mM Tris-HCl (pH7.4) 1mM EDTA, 1mM EGTA, 1% (v/v) NP-40.

Wash buffer II (WBII): WBI containing 0.5M NaCl.

Wash buffer III (WBIII): As WBI but containing NP-40 at 0.1% (v/v).

Prior to immunoprecipitation, the [^{35}S] Met-labelled

crude cell extract (~25µl) was pre-adsorbed against an equal volume of Protein A-sepharose beads (50% suspension in PBS) for one hour at room temperature, on a rotating platform. This step reduces non-specific binding of labelled proteins by Protein A sepharose in later stages of the procedure.

The Protein A-sepharose beads were pelleted by briefly spinning in a micro-centrifuge and discarded. The supernatant was transferred to a fresh tube and mixed with 20µl of antiserum. After two hours incubation on a rotating platform at room temperature, 100µl of fresh Protein A-sepharose beads were added and the incubation continued for sixty minutes more.

The beads and bound proteins were recovered by centrifugation for twenty seconds and washed sequentially in three changes each of WBI, WBII and WBIII. Proteins were then eluted from the beads by boiling for 3 minutes in SDS-sample loading buffer.

Labelled proteins were identified by SDS-PAGE and autoradiography of the dried gel as described in Sections 2.10 and 2.2.6 respectively.

2.10 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 pH8.8, 0.4% (w/v) SDS.
Stacking gel buffer:	0.5M Tris-HCl pH6.8, 0.4% (w/v) SDS.
AMPS:	10% (w/v) ammonium persulphate, made fresh daily.
Sample loading buffer:	50mM Tris-HCl pH6.8, 1% (w/v) SDS, 10% (v/v) glycerol, 0.05% (w/v) bromophenol blue, 1% v/v 2-mercapto-ethanol.

To prepare a 10% gel: 10ml acrylamide stock solution, 7.5ml resolving gel buffer, 50ul AMPS and dH_2O to 30 ml were mixed. 20µl TEMED were added and the resolving gel mixture poured into a 16 x 16 0.1cm mould formed by two glass plates, one notched at one end, one square, separated by 0.1cm spacers and sealed round the edges with a rubber gasket. The mould was filled to within 5cm of the top of the notched plate, then the gel

solution was carefully overlaid 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, 25ul 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 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 2.9.2 and 2.9.3 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 solution 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) methanol, 10% (v/v) glacial acetic acid over several hours at 37°C.

Once destained the gel was transfered to blotting paper and dried on a heated vacuum gel dryer. Gels containing [^{35}S]-labelled proteins were then autoradiographed as described in Section 2.2.6

2.11 COMPUTER METHODS

The University of Wisconsin Genetics Computer Group (GCG) sequence analysis software package, mounted on a VAX operating system, was used in most sequence analyses. The PROSEARCH programme (Collins *et al* 1988) run on a distributed array processor, mounted on a UNIX operating system, and the SERC SEQNET facility at Daresbury were used in protein sequence comparisons described in Chapter 3 below.

The MAP and Codon Composition programmes 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.

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.

2.12 DENSITOMETRY

Densitometry was performed using a Shimadzu dual-wavelength thin-layer chromato scanner. Linear scans were performed on low density autoradiograms using a single wavelength of 525nm, a slit width of 0.2mm, slit height of 3-6mm and a scan step of 0.1mm. Integration of the peaks was performed automatically by the scanner.

CHAPTER 3

3.1 Introduction

The abundance of eubacterial sigma factors has clearly established the role of alternative sigmas in gene regulation. Their ability to confer a new transcriptional specificity on RNAP provides an efficient means of co-ordinately activating unlinked genes, whilst the replacement of the primary sigma with an alternative sigma on a majority of the core enzyme molecules would also reduce expression of sigma 70-dependent genes.

As described in Chapter 1 at least 5 alternative sigmas co-ordinately regulate the expression of various stress-response genes in *E. coli*. The gene encoding one of these - sigma E - has not been identified, whilst further novel *E. coli* sigmas may yet be found. For example, the cellular response to cold shock and oxidative stress are global responses as yet not completely characterised, whilst other stress conditions which have not been mimicked in the laboratory may require additional sigmas.

Whether *E. coli* has further sigma genes is therefore a pertinent question relative to our understanding of the mechanisms governing bacterial gene expression.

All of the known *E. coli* alternative sigmas were discovered by workers investigating the positive regulation of *E. coli* genes

induced by a variety of stresses (see Section 1.8) of this thesis and references therein). However, this approach is limited by our knowledge of the types of stress to which bacteria may have to adapt physiologically. In a different approach, Fujita *et al* (1987) utilised the latest data then available on sigma sequence homology (Gribskov and Burgess 1986) in an attempt to identify new sigmas, by screening for proteins carrying antigenic epitopes similar to a conserved region of sigma. The authors raised polyclonal antisera against a synthetic tetradecameric peptide matching exactly the amino acid sequence of region 2.2, which is identical in sigma 70 and sigma 32 of *E. coli*. Indeed the latest comparative analysis by Lonetto *et al* (1992) found that region 2.2 is the most highly conserved region throughout eubacterial sigmas. The anti-peptide antisera specifically cross-reacted with both sigma 70 and sigma 32 on Western blots, and with about 10 other *E. coli* proteins, termed sigma cross reacting proteins (SCRPs) (Fujita *et al* 1987, Ueshima *et al* 1992). Four major SCRPs were subsequently purified and N-terminally sequenced and two of them - SCRP-27B and SCRP-34 - were identified as the ribosomal protein S2 and thioredoxin reductase. The remaining two - SCRP-23 and SCRP-27A - were not identifiable from their N-terminal amino acid sequences. Additionally none of the 4 SCRPs was found exclusively within the RNAP holoenzyme fraction upon glycerol gradient fractionation of *E. coli* crude lysates nor could they be shown to bind to core enzyme (Fujita *et al* 1987, Ueshima *et al* 1992).

In this chapter I describe the mapping and characterisation of the gene encoding SCRP-23. The protein, whose apparent molecular weight of 23kDa is implied by its name, seems to be present in significant amounts within exponentially growing cells and cross-reacts strongly with anti-2.2 peptide antibodies on Western blots. Its N-terminal sequence is SLINTKIKPKNQAFKNGEFIEITEKDTE (Ueshima *et al* 1992). Whilst SCRP-23 appears to be more abundant than known *E. coli* sigmas, and failed to associate with RNAP core enzyme under various conditions, Fujita *et al* (1987) argued that certain sigmas might only associate with RNAP through interaction with the DNA template.

In order to determine the nature and identity of SCRP-23, I decided to clone and sequence the gene encoding it (provisionally named *scrp23*). Meanwhile further biochemical characterisation of SCRP-23 was carried out by Ishihama and colleagues at the National Institute of Genetics (NIG), Mishima, Japan (Ueshima *et al* 1992).

3.2.1 Location of the Putative Scrp23 Gene on the Physical Map

Based on the N-terminal aa sequence of SCRP-23 a pair of of degenerate oligonucleotide probes, Scrp23-1 [5'-CA(G/A)GCNTT(T/C)AA(G/A)AA(T/C)GC-3'] corresponding to aa 13 through 18 and Scrp23-2 [5'-GA(G/A)AA(G/A)GA(T/C)ACNGA(G/A)GG-3'] corresponding

to aa 25 through 30, were designed by N. Fujita, N.I.G., Japan. Both probes were 5' labelled using gamma-[^{32}P] ATP as described in Section 2.8.2 and hybridised with nylon filters carrying DNA from the miniset lambda clones of the *E. coli* K-12 genome (Kohara *et al* 1987, Noda *et al* 1991) by Dr. R. S. Hayward (then working at N.I.G.) as described in Section 2.8.5.

Scrp23-1 hybridised strongly with a pair of contiguous clones, lambda 166 and lambda 167 (data not shown), whose *E. coli* DNA inserts overlap by approximately 9.8 kb and are located near 650 kb on the physical map, (14 minutes on the genetic map) of the *E. coli* K-12 chromosome (Kohara *et al* 1987, Bachmann 1990). Scrp23-2 hybridised weakly to a number of clones in repeated hybridisations (data not shown). Of these, the three contiguous lambda clones 128, 129 and 130 were the most prominent. There is however no common overlap between these three clones, which map near 259 kb on the chromosome. Apparently scrp23-2 hybridised weakly with at least two sites within this region. Scrp23-2 did not hybridise with lambda 166 or lambda 167.

Despite the inconsistent hybridisation patterns shown by scrp23-1 and scrp23-2, the stronger and more specific pattern produced by scrp23-1 suggested lambda 166 and lambda 167 as the likeliest candidate *scrp23* clones.

3.2.2 Fine Mappping of the Putative *Scrp23* Gene

The location of the *scrp23* gene on the physical map was confirmed and more accurately defined by restriction mapping and Southern hybridisation analysis of lambda 166 and lambda 167 DNA using [^{32}P] labelled scrp23-1. DNA was prepared from concentrated lysates of lambda 166 and lambda 167 (previously amplified on the *recD* strain NM621, which is 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 (figure 3.1) confirmed the identity of lambda 166 and lambda 167, and gave results largely consistent with the consensus map of the *E. coli* K-12 chromosome (Kohara *et al* 1987), apart from a few minor differences. Scrp23-1 hybridised to a region within the overlap between lambda 166 and lambda 167, flanked by *BglI* sites near to 653 kb and 654 kb respectively on the physical map (figure 3.2), and therefore near 14.2 minutes on the genetic map (Kohara *et al* 1991, Bachman 1990). The closest mapped genes on either side of this region are *entA*, located near 645 kb/13.85 minutes (Nahlik *et al* 1987), and *rna* at 659 kb/14.4 minutes (Meador III and Kendal 1990). The above data are summarised in figure 3.3.

Figure 3.1

Restriction analysis of lambda 166 and 167 DNAs. 0.6% agarose gel electrophoresis. Tracks 2, 6, 10 and 12 contain lambda 166 DNA; tracks 3, 5, 7, 9, 11 and 13 contain lambda 167 DNA. Tracks 1 and 14 show lambda *c*I857 *S*am7 marker DNA digested with *Hin*dIII and *Hin*dIII + *Eco*RI respectively. The relevant marker bands are marked and the sizes given (in base pairs) at the sides. Tracks 4 and 8 are empty. The following abbreviations are used for restriction enzymes: E, *Eco*RV; G, *Bgl*I; H, *Hin*dIII; K, *Kpn*I and R, *Eco*RI.

FIGURE 3.1

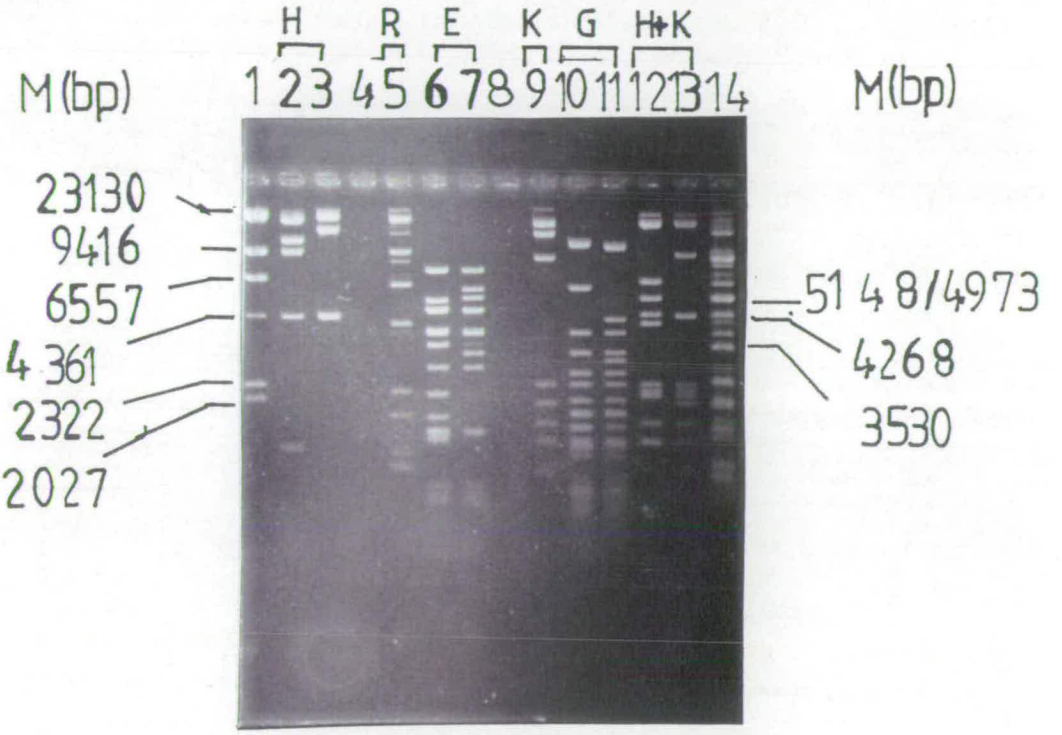

H
R
E
K
G
H+K
M(bp)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
M(bp)
23130
9416
6557
4361
2322
2027
5148/4973
4268
3530

Figure 3.2

Fine mapping of the putative *scrp23* gene by Southern blotting using 5'-[^{32}P]-labelled Scrp23-1 as probe. DNA was transfered to a nylon membrane from the gel shown in Figure 3.1 as described in materials and methods. Hybridisation was performed at 37°C for 16 hours. Final washing conditions were 6 x SSC, 0.2% SDS at 30°C for two periods of 15 minutes. More stringent washes were not required in this case. Tracks 2, 6, 10 and 12 contain lambda 166 DNA; Tracks 3, 5, 7, 9, 11 and 13 contain lambda 167 DNA; Tracks 4 and 8 are empty, and tracks 1 and 14 contain lambda *c*I857 *S*am7 marker DNA digested with *Hin*dIII and *Hin*dIII + *Eco*RI respectively; the relevant bands are marked and their sizes given (in base pairs) at the sides. Filled arrow-heads mark the positions of probe positive *Bgl*I fragments (tracks 10 and 11) which define the limits of the positively hybridising region, and a probe positive ~4.1 kilobase *Hin*dIII - *Kpn*I fragment (track 12) selected for sub-cloning of the putative *scrp*23 gene. The following abbreviations are used for restriction enzymes: E, *Eco*RV; G, *Bgl*I; H, *Hin*dIII; K, *Kpn*I and R, *Eco*RI.

FIGURE 3.2

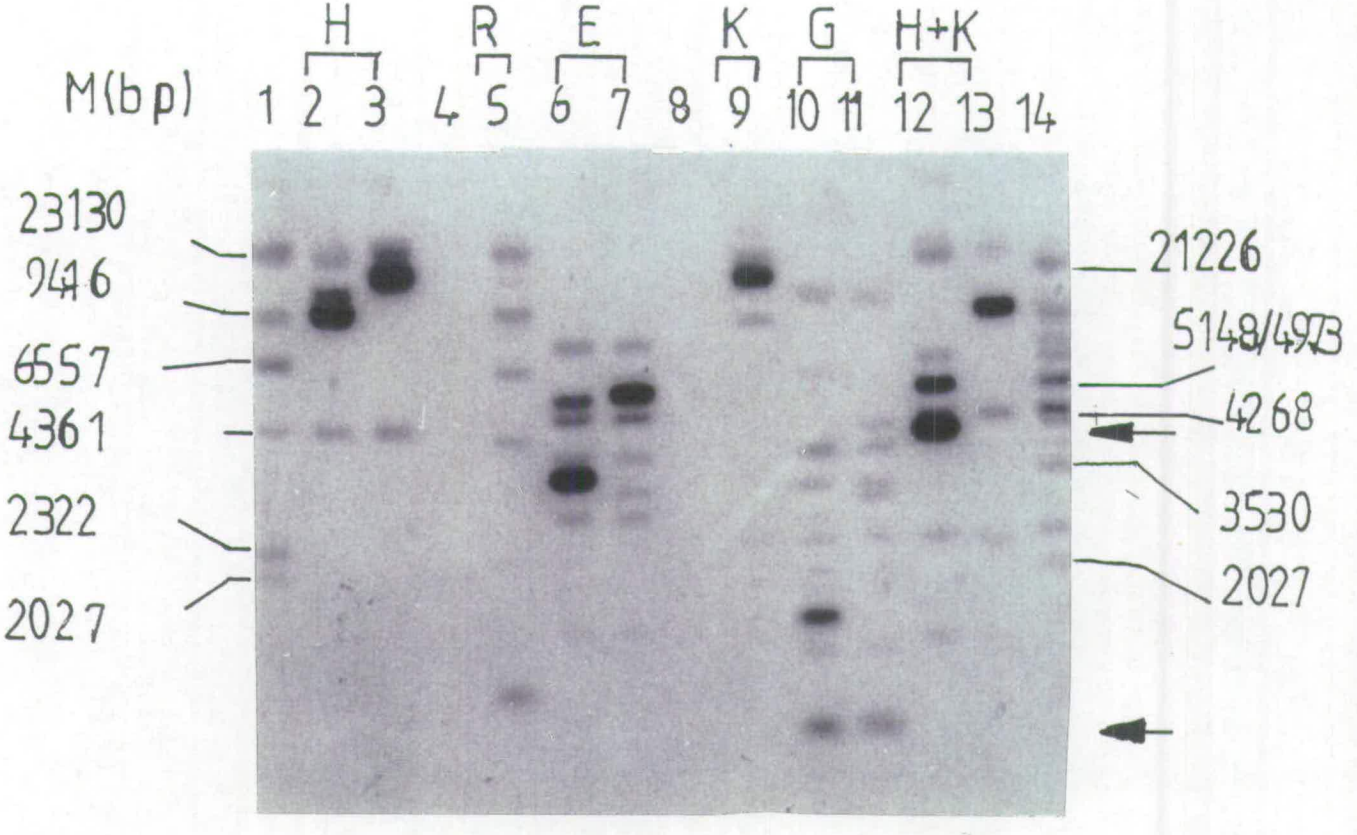
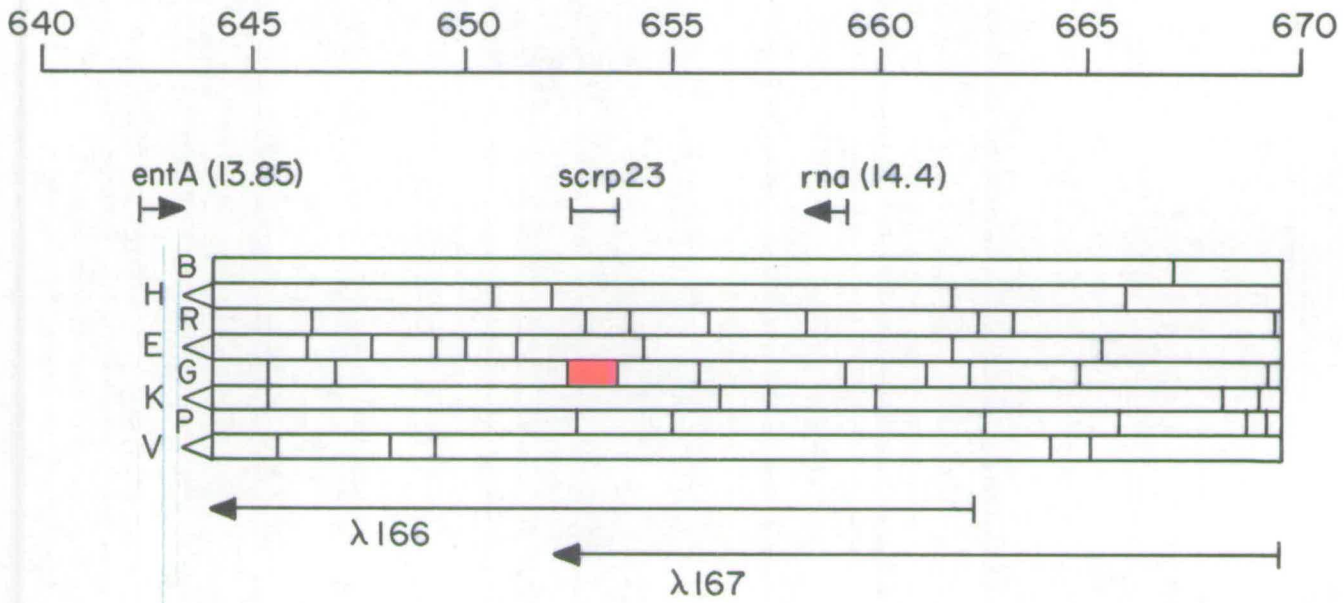

H
R
E
K
G
H+K
M(bp)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
23130
9416
6557
4361
2322
2027
21226
5148/4973
4268
3530
2027

Figure 3.3

Restriction map of *E. coli* chromosomal DNA inserts cloned in lambda 166 and lambda 167. This map is based on restriction analyses and the physical map of Kohara *et al* (1987). The top scale shows co-ordinates on the physical map (*ibid*) in kilobases. The locations of the *ent*A *scrp*23 and *rna* genes are marked; arrows indicate the gene orientation where known. The ~1.04 kilobase *Bgl*I fragment which hybridised with scrp23-1 is highlighted in red. The length and orientation of the chromosomal DNA inserts carried by lambdas 166 and 167 are indicated below the restriction map. The following abbreviations are used for restriction enzymes: B, *Bam*HI; E, *Eco*RV; G, *Bgl*I; H, *Hind*III; K, *Kpn*I; P, *Pst*I and V, *Pvu*II.

FIGURE 3.3


640
645
650
655
660
665
670
entA (13.85)
scrp23
rna (14.4)
B
H
R
E
G
K
P
V
λ166
λ167

3.3 Subcloning the Putative *Scrp23* Gene

Southern hybridisation analysis located the upstream end of the putative *scrp23* gene within a 1.04 kb *Bgl*I fragment within the overlap between lambda 166 and lambda 167. As neither the precise location nor the orientation of the putative *scrp23* gene with respect to the *Bgl*I sites was known, a larger (4.1 kb) *Hin*dIII-*Kpn*I fragment which overlaps either end of the *Bgl*I fragment was subcloned between the corresponding sites in M13 mp19 (figure 3.4).

The structure of the resulting recombinant, M13 mp19-22, was verified by restriction mapping (data not shown).

A plasmid subclone of the *scrp23* locus was then constructed by inserting the 4.8 kb *Hin*dIII - *Bgl*II fragment from mp19-22 between the *Hin*dIII and *Bam*HI sites of pACYC184. The resulting Cm^r, Tet^s recombinant - pDSA23 -contains the entire *Hin*dIII-*Kpn*I fragment of *E. coli* W3110 genomic DNA, together with a 692 bp linker fragment of M13-mp19 DNA stretching from the *Kpn*I target within the multiple cloning sequence to the *Bgl*II site within the M13 *gene II*. Restriction analysis of pDSA23 (figure 3.5) revealed an additional *Pst*I site, in the *E. coli* DNA, not recorded by Kohara *et al* (1987). It lies near 653.2 kb on the consensus map, between the known *Pst*I (652.8 kb) and *Eco*RI (653.5 kb) sites. This site is located only 0.4 kb away from the known *Pst*I target and was probably overlooked during large scale mapping, although the possibility that it was generated spontaneously by mutation, for example during

amplification of pDSA23, cannot be discounted.

Figure 3.4

Summary of the construction of mp19-22 (not to scale). The red-filled segments in the M13 DNA outlines represent the *lac Z'* gene fragment and the gene II of the vector. The hatched rectangles represent the probe positive chromosomal fragment of lambda 166. The following abbreviations are used for restriction enzymes: H, *Hin*dIII; L, *Bgl*II and K, *Kpn*I.

FIGURE 34

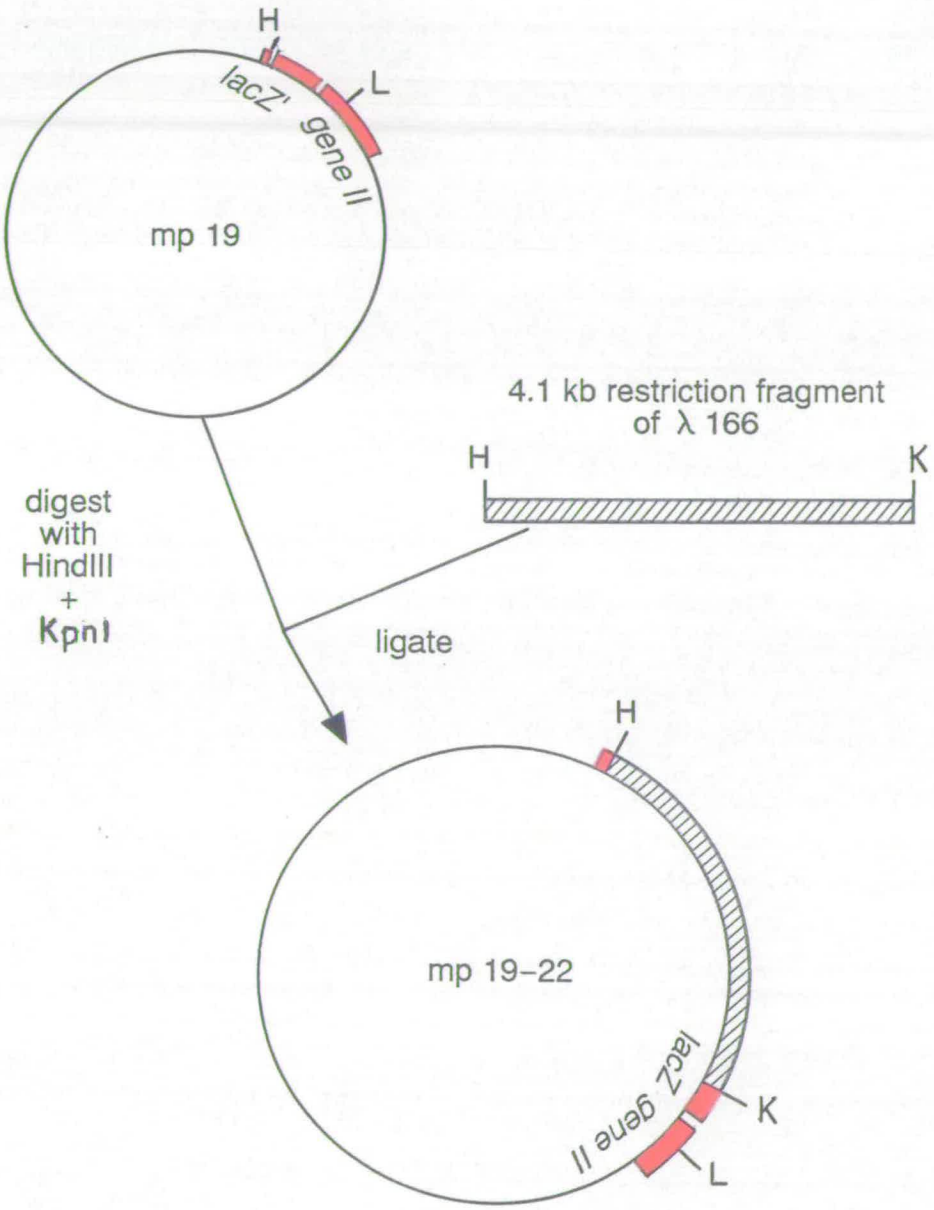
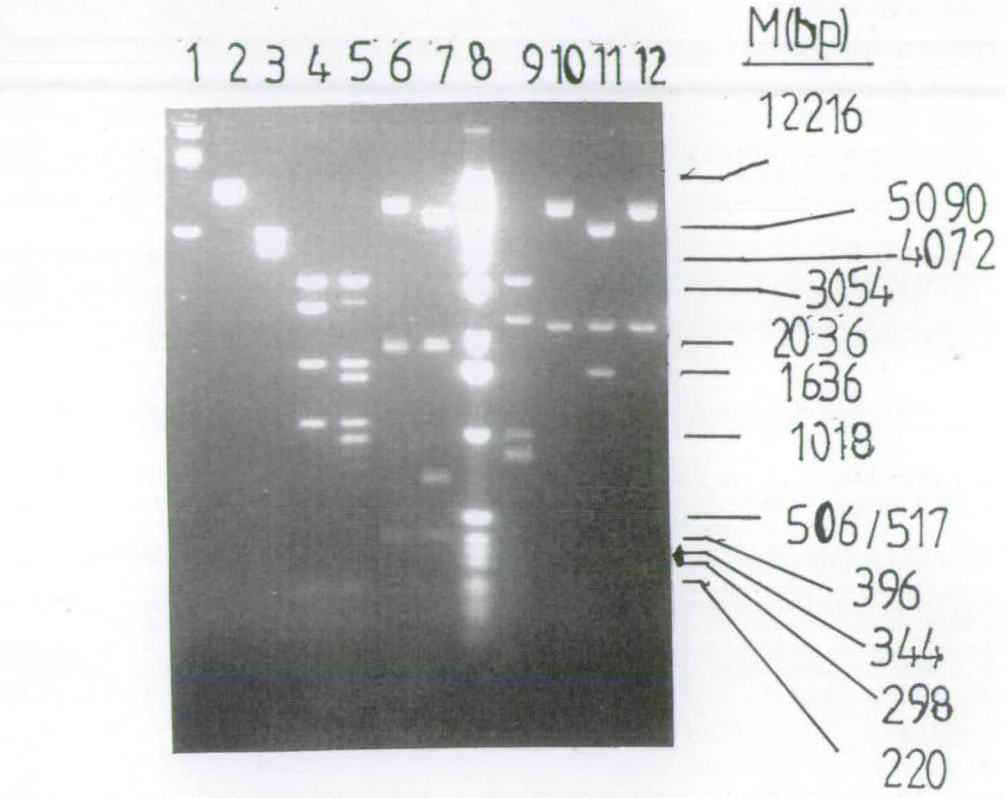
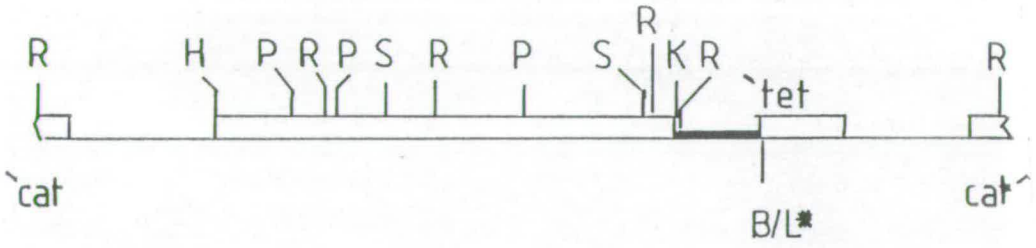

H
lacZ'
L
gene II
mp 19
4.1 kb restriction fragment
of λ 166
H
K
digest
with
HindIII
+
KpnI
ligate
H
mp 19–22
lacZ gene II
K
L

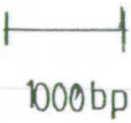
Figure 3.5

Restriction analysis of pDSA23. 1% agarose gel electrophoresis. The top part shows the restriction gel. Track 1 shows uncut pDSA23; Track 8 shows 1 kilobase ladder marker DNA (Gibco BRL), relevant fragments are marked and their sizes given (in base pairs) at the side. The remainder of the tracks show pDSA23 digested with the following enzymes: Track 2, *Hin*dIII; Track 3, *Hin*dIII + *Kpn*I; Track 4, *Eco*RI; Track 5, *Hin*dIII + *Eco*RI; Track 6, *Pst*I; Track 7, *Hin*dIII + *Pst*I; Track 9, *Eco*RI + *Pst*I; Track 10, *Sma*I; Track 11, *Sma*I + *Hin*dIII and Track 12, *Sma*I + *Kpn*I. The bottom part shows a restriction map of pDSA23, linearised at the *Eco*RI site within the *cat* gene (position 1 in pACYC184). Open rectangles in the plasmid DNA outline represent the *'cat*, *cat'* and *tet* gene fragments and the ~4.1 kilobase *E. coli* genomic DNA insert. The thick black line represents the 692 base pair linker fragment of mp19 DNA. *cat* encodes chloramphenicol acetyl transferase and *tet* encodes tetracycline resistance. The following abbreviations are used for restriction enzymes: H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; R, *Eco*RI and S, *Sma*I. B/L denotes that *Bam*HI(B) and *Bgl*II(L) ends were ligated together thus preventing the regeneration of either site.

FIGURE 3.5


1 2 3 4 5 6 7 8 9 10 11 12
M(bp)
12216
5090
4072
3054
2036
1636
1018
506/517
396
344
298
220


R
H
P
R
P
S
R
P
S
R
K
R
R
`tet
`cat
B/L*
cat´


1000bp

3.4 In Vivo Expression of SCRP-23

The strong cross-reaction of SCRP-23 during Western blot analysis and its purification in significant amounts from *E. coli* cells suggest that it is an abundant protein (Fujita *et al* 1987, Ueshima *et al* 1992). Proteins encoded by the cloned inserts in lambda 166 and lambda 167 and the *Hin*dIII-*Kpn*I insert DNA of pDSA23 were characterised by examining protein synthesis *in vivo*, using UV irradiated host cells and purified minicells respectively. The *scrp23* gene product was identified by selective immunoprecipitation of labelled plasmid-encoded proteins from minicell extracts.

3.4.1 Analysis of Proteins Encoded by Lambda 166 and Lambda 167

S159 or S159 (lambda *ind*) cells were UV-irradiated and infected with lambda 166, lambda 167 or one of three control phages - lambda 440, lambda 639 or lambda 640 - from the miniset collection, as described in Section 2.9.3. The cells were then pulse-labelled 10-13 minutes post infection with [^{35}S]Met and protein samples fractionated by SDS-Polyacrylamide Gel electrophoresis (SDS-PAGE) as described in Section 2.10.

As figure 3.6 shows, polypeptides of estimated MW 53kDa and 22.5kDa were strongly synthesised only when S159 and S159 (lambda *ind*) cells were infected with lambda 166 or lambda 167 tracks. Proteins of similar MW were not

observed in uninfected control cells or cells infected with control phages. Synthesis of both polypeptides appeared to be stronger in S159 (lambda *ind*) cells, especially for lambda 166, and the 22.5kDa polypeptide appeared more strongly synthesised than the 53kDa protein in all cases.

The apparent abundance of both proteins in the lysogenic host S159 (lambda *ind*) suggests that the genes encoding the 22.5kDa and 53kDa polypeptides are strongly expressed from autogenous promoter(s). The estimated MW of the 22.5kDa protein and its apparent abundance correlate well with the observation of Fujita *et al* (1987) that the sigma cross-reacting protein of 23kDa is an abundant protein.

3.4.2 Analysis of Proteins Encoded by pDSA-23

Polypeptides produced by the *Hin*dIII-*Kpn*I insert of pDSA23 were observed by [^{35}S]Met labelling minicells.

Minicells were purified from cultures of DS410 harbouring pDSA23 or pACYC184 (control) and labelled with [^{35}S]Met as described in Sections 2.9.1 and 2.9.2. Non-transformed minicells, isolated and labelled in the same way, provided negative control samples. When crude cell lysates were fractionated by SDS-PAGE and autoradiographed, proteins of estimated MW 53kDa and

22.5kDa were synthesised specifically in cells transformed by pDSA23 (figure 3.7a). The 22.5kDa protein is partially obscured in figure 3.7a (track 2) by the vector encoded CAT gene product (compare track 1). An additional polypeptide of estimated MW 17kDA was encoded by pDSA23, which was not observed in UV irradiated cells infected with lambda 166 or 167. This polypeptide was strongly synthesised and may be a hybrid protein resulting from fusion of insert and vector DNA sequences during the construction of pDSA23.

3.4.3 Immunoprecipitation of SCRP-23 from Minicell Extracts

The remaining protein samples prepared in Section 3.4.2. were immunoprecipitated with the antipeptide antiserum used in the original isolation of SCRP-23 (Fujita *et al* 1987), as described in Section 2.11.

Precipitates were then fractionated by SDS-PAGE and radiolabelled proteins observed by autoradiography. As Figure 3.7b (track 2) shows, the 22.5kDa protein encoded by the *HindIII-KpnI* insert of pDSA23 was selectively precipitated by the antipeptide antiserum, suggesting that the gene encoding SCRP-23 is indeed situated within the cloned region of pDSA23. Intriguingly, the 53kDa polypeptide appears to have been co-precipitated with SCRP-23 protein from pDSA23 extracts. The absence of the highly expressed chloramphenicol acetyl transferase

(CAT) protein from the immunoprecipitate (track 3) corroborates this observation, suggesting that the 22.5kDa and 53kDa polypeptides may be able to form complexes, and could be functionally related. However an alternative explanation, that the 53kDa protein cross-reacts with another species of antibody, cannot be discounted. However it should be noted that no SCRP of 53kDa was detected by Ishihama and co-workers (Fujita *et al* 1987).

Figure 3.6

Detection of ^{35}S-labelled-proteins synthesised by lambdas 166 and 167 in UV irradiated *E. coli* by autoradiography of a 10% polyacrylamide gel. Tracks 1, 3, 5, 7, 9 and 11 contain S159 crude extracts, Tracks 2, 4, 6, 8, 10 and 12 contain S159 (lambda *ind*) crude extracts. Extracts prepared from cells infected by lambdas 166, 167, 440, 639 or 640 are indicated above the photograph. Tracks 1 and 8 show crude extracts from uninfected control samples. Proteins specifically encoded by lambda 166, lambda 167 or lambda 440 are marked on the left and their sizes given. The sizes of molecular mass standards (included in Track 8) are indicated on the right in kiloDaltons.

FIGURE 3.6

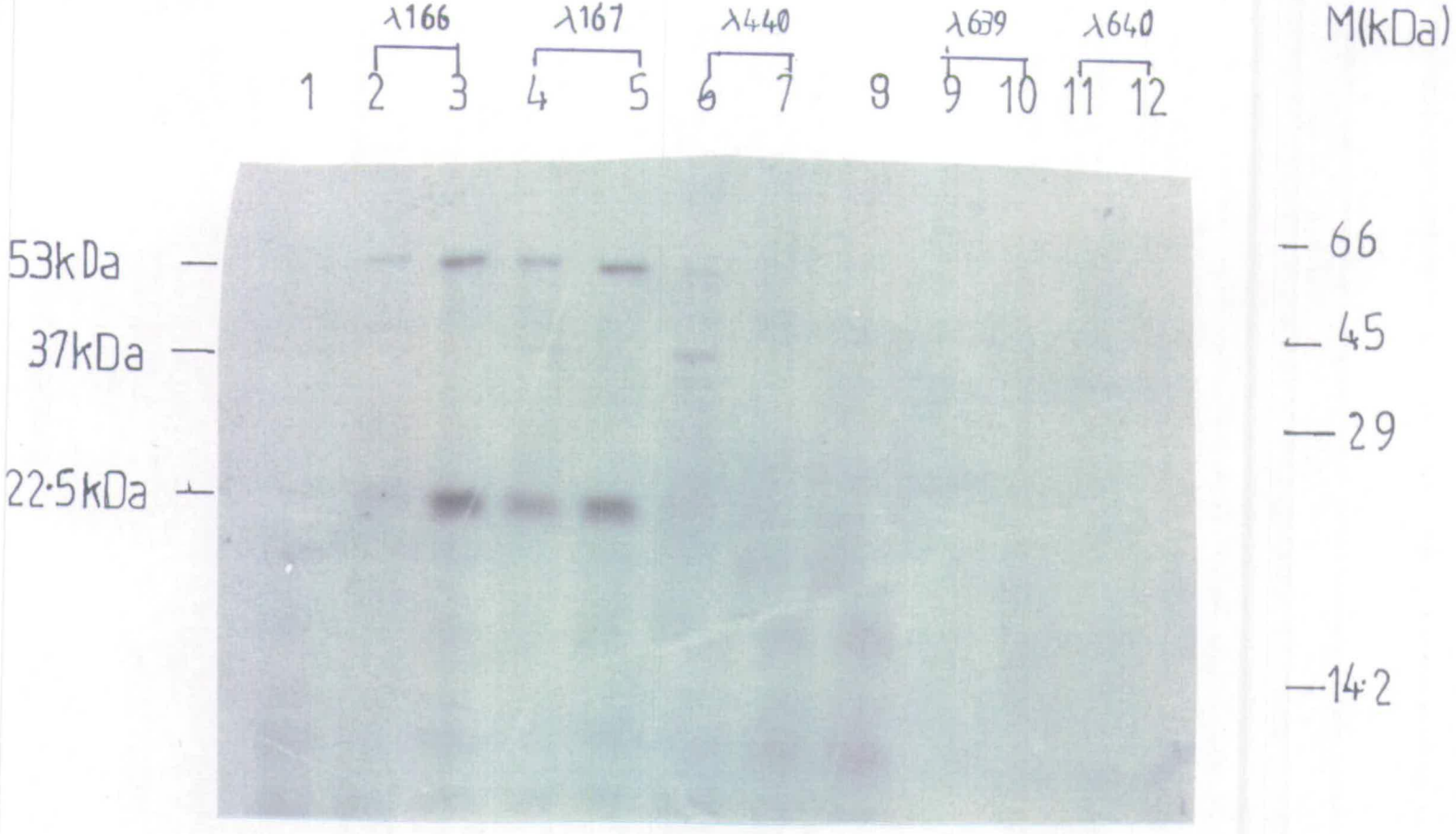

λ166
λ167
λ440
λ639
λ640
M(kDa)
1 2 3 4 5 6 7 9 9 10 11 12
53kDa
37kDa
22·5kDa
66
45
29
14·2

Figure 3.7

Proteins encoded by pDSA23 in minicells.

A. autoradiography of a 10% polyacrylamide gel. Tracks 1-3 show gene products in minicell extracts of strain DS410 bearing plasmids:

1) pACYC 184; 2) pDSA23 and 3) "No plasmid".

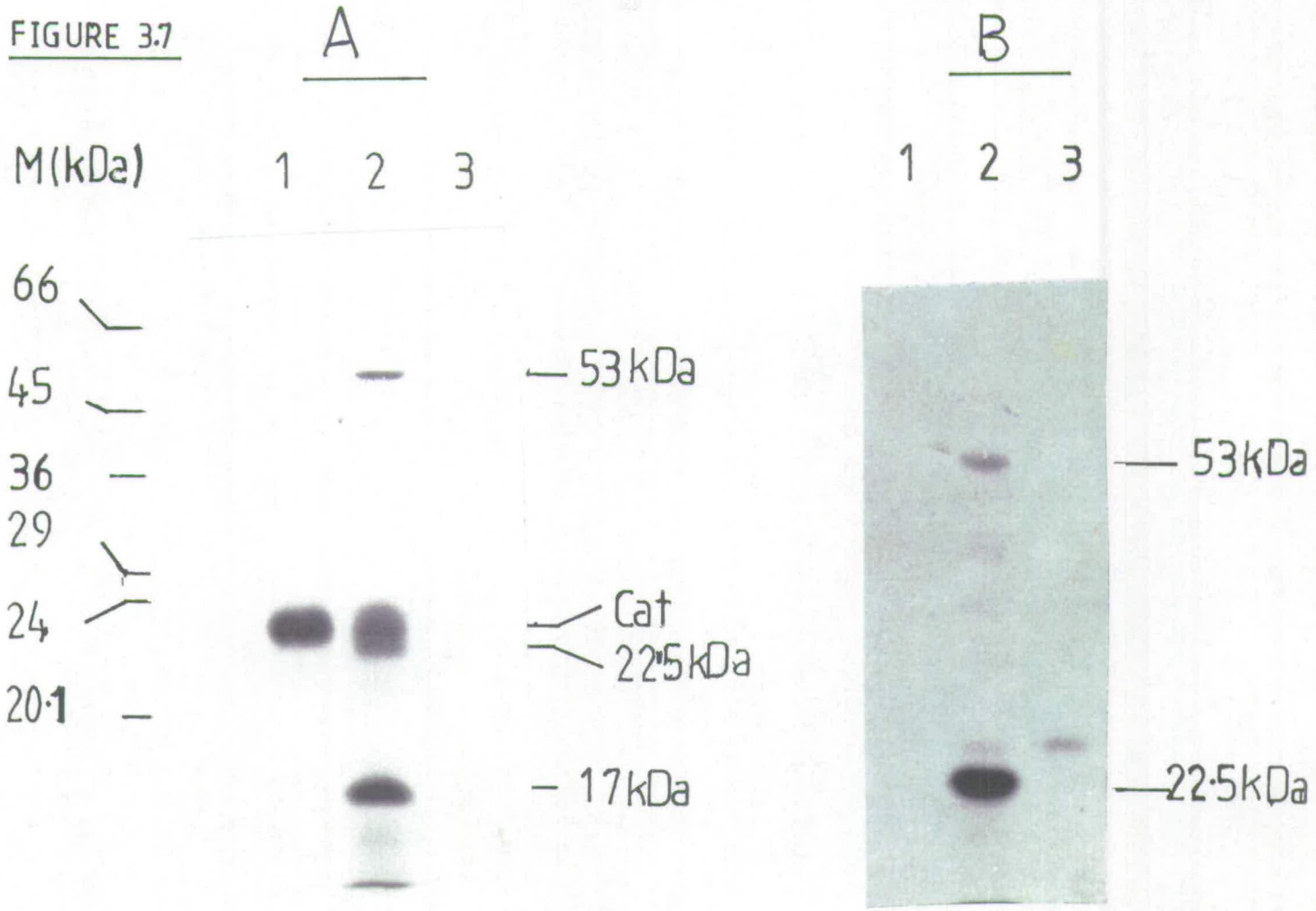
Proteins encoded by pDSA23 are marked and their sizes given on the right. CAT denotes the chloramphenicol acetyl transferase protein which is encoded by the vector. The sizes of molecular mass standards are indicated on the left.

B. SDS-polyacrylamide gel electrophoretic analysis of proteins immunoprecipitated from minicell extracts. Tracks 1-3 show proteins immunoprecipitated from DS410 minicell extracts harbouring plasmids:

1) No plasmid; 2) pDSA23 and 3) pACYC184.

Immunoprecipitated proteins are marked and their sizes shown on the right.

FIGURE 3.7


A
B
M(kDa)
1
2
3
1
2
3
66
45
36
29
24
20·1
53kDa
Cat
22·5kDa
17kDa
53kDa
22·5kDa

3.5 Nucleotide Sequence of the *scrp23* Gene

Having established that a SCRP of 23kDa was indeed synthesized from the *Hin*dIII-*Kpn*I insert of pDSA23, the nucleotide sequence of the *scrp23* gene was determined.

3.5.1 Construction of Subclones Covering the *scrp23* Region

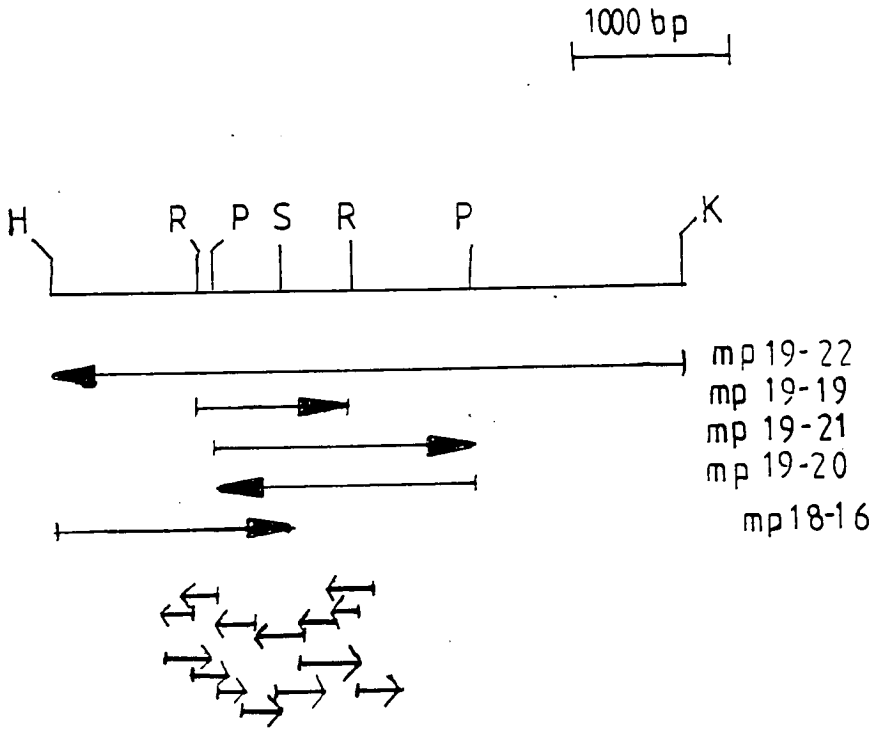
The *Hin*dIII-*Kpn*I insert of M13-mp19-22 is over 4kb in length. In order to facilitate sequencing of both strands of the *scrp23* gene a set of smaller overlapping fragments was subcloned into M13 mp18 and M13 mp19.

The ~ 1.3 kb *Hin*dIII-*Sma*I fragment from pDSA23 was cloned between the same sites in M13 mp18 to produce M13 mp18-16. 1.04 kb *Eco*RI and 1.8 kb *Pst*I fragments of M13 mp19-22 were cloned into the corresponding sites of M13 mp19. The size and orientation of each fragment in the subclones was verified by restriction analysis (data not shown). Subclones containing the 1.04 kb *Eco*RI fragment were isolated in only one orientation, as judged by digestion with *Pst*I. One of these recombinants was named M13 mp19-19. Subclones were isolated containing the *Pst*I fragment in either orientation with respect to the *Eco*RI site in the polylinker of M13 mp19. These were named M13 mp19-20 and M13 mp19-21. A restriction map showing the inserts in mp18-16, mp19-19, mp19-20 and mp19-21 is given in figure 3.8.

Figure 3.8

M13 sub-clones covering the *scrp23* locus; and sequencing strategy. The top line shows a restriction map of the *E. coli* chromosomal DNA insert in mp19-22. The extent and orientation of DNA fragments cloned into M13 mp18 or M13 mp19 are indicated by large arrows. Small arrows show the direction and approximate extent of sequencing reactions primed with different oligonucleotides. The following abbreviations are used for restriction enzymes: H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; R, *Eco*RI and S, *Sma*I.

FIGURE 3.8



3.5.2 Determination of the *scrp 23* Nucleotide Sequence

Nucleotide sequencing was initiated from the universal sequencing primer [5'-GTAAAACGACGGCCAGT-3'] using single stranded M13 mp19-19 DNA and M13 mp19-21 DNA, as described in Section 2.7.1. Further DNA sequence information was then generated using synthetic oligonucleotide primers complementary to already determined nucleotide sequences. The sequencing strategy is presented in figure 3.8

In 1483 nucleotides of DNA sequence (figure 3.9), two long open reading frames, separated by 246 base pairs, were identified. Both open reading frames (upon comparison with the restriction map of the *scrp*23 locus) are predicted to be transcribed in the same, clockwise direction on the *E. coli* chromosome (in parallel with DNA replication of the region). The more upstream reading frame extends for 187 codons (excluding the TAA stop codon) and is preceded by a convincing Shine-Dalgarno consensus ribosome binding sequence (underlined in figure 3.9). The downstream open reading frame extends for at least 162 codons, and again is preceded by a good Shine-Dalgarno consensus element just upstream of the initiating ATG codon. Two potential sigma 70 promoter elements were identified in the non-coding region preceding the upstream open reading frame. The -10 and -35 consensus sequences of both potential

promoters are boxed in figure 3.9. Additionally, no likely promoter or transcription termination signals were detected within the intergenic DNA sequence.

Comparison of the predicted amino acid sequences near the beginning of both open reading frames with the N-terminal amino acid sequence determined for SCRP23 (Ueshima *et al* 1992) identified the upstream reading frame as the gene encoding SCRP23.

3.5.3 Amino Acid Composition of SCRP23 and Codon Usage in *scrp23*

The amino acid composition of SCRP23 and the codon usage of the *scrp23* gene were analysed using the pepsort and codon frequency programmes respectively of the CGC sequence analysis software package

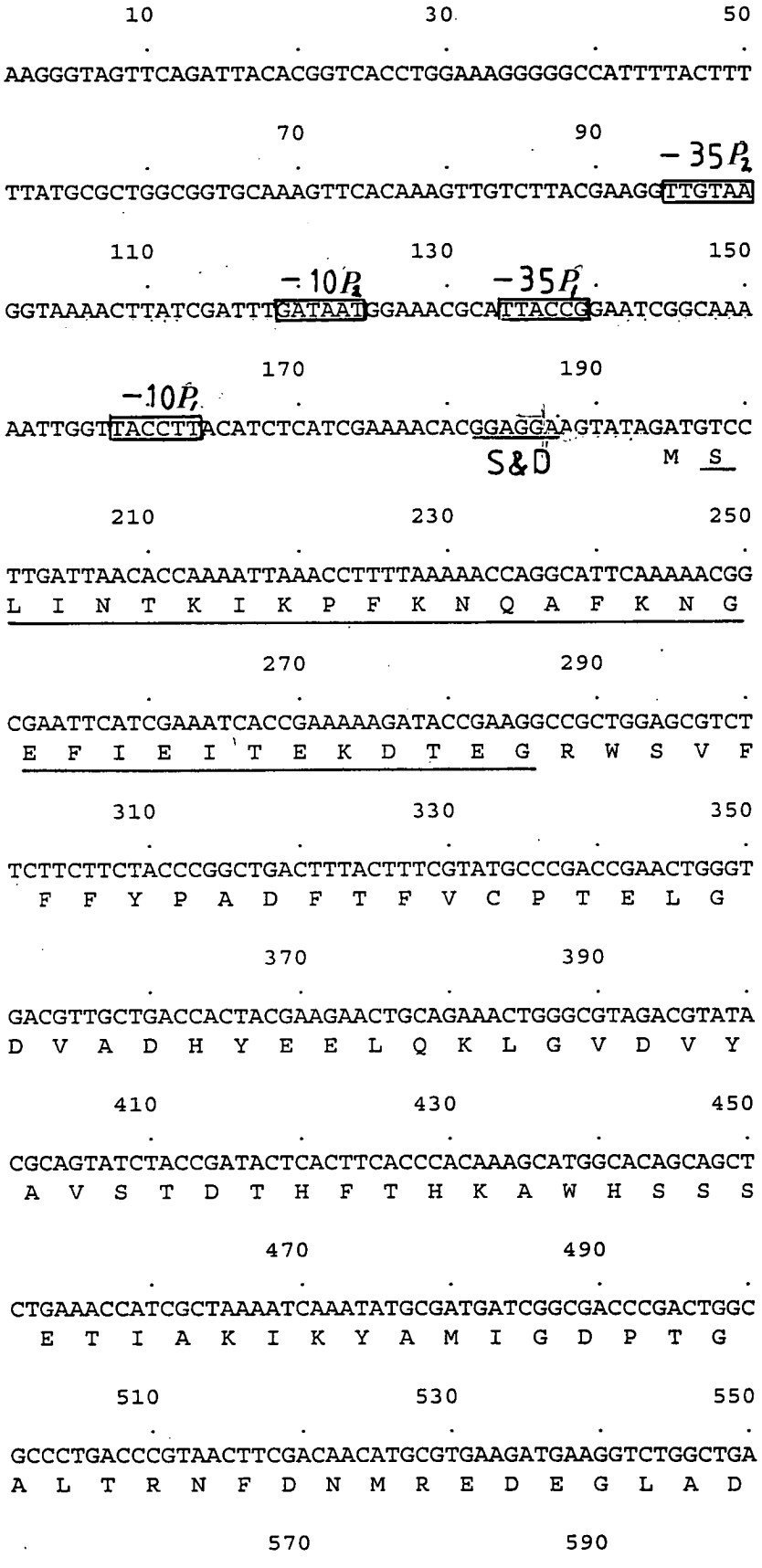
The amino acid composition of SCRP23, as deduced from the *scrp23* nucleotide sequence, is summarised in Table 3.1. The predicted molecular weight (MW) of 20,846 is in good agreement with the 22,500 value estimated from SDS-PAGE analysis in Sections 3.4.1 and 3.4.2.

The codon usage of the *scrp23* gene is summarised in Table 3.2. It was found to be non-random, showing a strong positive correlation between codon choice and major iso-accepting tRNA species of *E. coli* (Ikemura 1981 and 1985). The frequency of optimal codons (Fop) 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 high Fop of 0.91 is consistent with the observation that *scrp*23 is highly expressed.

Figure 3.9

Nucleotide sequence of the *scrp*23 gene and predicted aa sequence of the encoded protein. The nucleotide sequence of the *scrp*23 gene, its flanking regions and a downstream open reading frame is shown in the 5′ to 3′ direction. The predicted aa sequences of the SCRP-23 protein and the translation product of the downstream *orf* are shown below the relevant codons. The underlined aa sequence is identical to the N-terminal aa sequence of SCRP-23 reported by Ueshima *et al* (1992). The likely Shine and Dalgarno sequences upstream of *scrp*23 and the downstream *orf* are underlined, and the -10 and -35 elements of the two putative promoters upstream of *scrp*23 are indicated.

The sequence was submitted to the DNA Database of Japan (DDBJ) and was given the accession number D13187.


10
30
50
AAGGGTAGTTCAGATTACACGGTCACCTGGAAAGGGGGCCATTTTACTTT
70
90
−35 P2
TTATGCGCTGGCGGTGCAAAGTTCACAAAGTTGTCTTACGAAGGTTGTAA
110
130
150
−10 P2
−35 P1
GGTAAAACTTATCGATTTGATAATGGAAACGCATTACCGGAATCGGCAAA
170
190
−10 P1
AATTGGTTACCTTACATCTCATCGAAAACACGGAGGAAGTATAGATGTCC
S&D
M S
210
230
250
TTGATTAACACCAAAATTAAACCTTTTAAAAACCAGGCATTCAAAAACGG
L I N T K I K P F K N Q A F K N G
270
290
CGAATTCATCGAAATCACCGAAAAAGATACCGAAGGCCGCTGGAGCGTCT
E F I E I T E K D T E G R W S V F
310
330
350
TCTTCTTCTACCCGGCTGACTTTACTTTCGTATGCCCGACCGAACTGGGT
F F Y P A D F T F V C P T E L G
370
390
GACGTTGCTGACCACTACGAAGAACTGCAGAAACTGGGCGTAGACGTATA
D V A D H Y E E L Q K L G V D V Y
410
430
450
CGCAGTATCTACCGATACTCACTTCACCCACAAAGCATGGCACAGCAGCT
A V S T D T H F T H K A W H S S S
470
490
CTGAAACCATCGCTAAAATCAAATATGCGATGATCGGCGACCCGACTGGC
E T I A K I K Y A M I G D P T G
510
530
550
GCCCTGACCCGTAACTTCGACAACATGCGTGAAGATGAAGGTCTGGCTGA
A L T R N F D N M R E D E G L A D
570
590

```
0       .         .         .         .         .
CCGTGCGACCTTCGTTGTTGACCCGCAGGGTATCATCCAGGCAATCGAAG
 R  A  T  F  V  V  D  P  Q  G  I  I  Q  A  I  E  V

      610                 630                 650
        .         .         .         .         .
TTACCGCTGAAGGCATTGGCCGTGACGCGTCTGACCTGCTGCGTAAAATC
  T  A  E  G  I  G  R  D  A  S  D  L  L  R  K  I

                670                 690
        .         .         .         .         .
AAAGCAGCACAGTACGTAGCTTCTCACCCAGGTGAAGTTTGCCCGGCTAA
K  A  A  Q  Y  V  A  S  H  P  G  E  V  C  P  A  K

      710                 730                 750
        .     0   .         .         .         .
ATGGAAAGAAGGTGAAGCAACTCTGGCTCCGTCTCTGGACCTGGTTGGTA
 W  K  E  G  E  A  T  L  A  P  S  L  D  L  V  G  K

                770                 790
        .         .         .         .         .
AAATCTAAATTTCCTTAGTCTTTCACGCATAGCGGCGTTGCGTCGCCCGC
  I

      810                 830                 850
        .         .         .         .         .
TCACCCGGTCACTTACTTGTGTAAGCTCCCGGGGATTCACAGCTAGCGCC

                870                 890
        .         .         .         .         .
TTGCTCTGACGCGAAATACTTCGGAAATTCACCTAATTCTTCGGGTGCTG

      910                 930                 950
        .         .         .         .         .
CGGCGCATTTTCTTCCCCGCACCATGATGCAAGCTGCATCCAGGTAGCCG

                970                 990
        .         .         .         .         .
CAGAGGCCGCTTGCATGATGATGTTTAAGAGCCCAGGAGATAAACATGCT
                                  S&D         M  L

     1010                1030                1050
        .         .         .         .         .
CGACACAAATATGAAAACTCAACTCAAGGCTTACCTTGAGAAATTGACCA
 D  T  N  M  K  T  Q  L  K  A  Y  L  E  K  L  T  K

               1070                1090
        .         .         .         .         .
AGCCTGTTGAGTTAATTGCCACGCTGGATGACAGCGCTAAATCGGCAGAA
  P  V  E  L  I  A  T  L  D  D  S  A  K  S  A  E

     1110                1130                1150
        .         .         .         .         .
ATCAAGGAACTGTTGGCTGAAATCGCAGAACTGTCAGACAAAGTCACCTT
I  K  E  L  L  A  E  I  A  E  L  S  D  K  V  T  F
```

```
                1170                1190
          .         .         .         .         .
TAAAGAAGATAACAGCTTGCCGGTGCGTAAGCCGTCTTTCCTGATCACCA
 K  E  D  N  S  L  P  V  R  K  P  S  F  L  I  T  N

      1210                1230                1250
          .         .         .         .         .
ACCCAGGTTCCAACCAGGGGCCACGTTTTGCAGGCTCTCCGCTGGGCCAC
  P  G  S  N  Q  G  P  R  F  A  G  S  P  L  G  H

                1270                1290
          .         .         .         .         .
GAGTTCACCTCGCTGGTACTGGCGTTGCTGTGGACCGGTGGTCATCCGTC
E  F  T  S  L  V  L  A  L  L  W  T  G  G  H  P  S

      1310                1330                1350
          .         .         .         .         .
GAAAGAAGCGCAGTCTCTGCTGGAGCAGATTCGCCATATTGACGGTGATT
 K  E  A  Q  S  L  L  E  Q  I  R  H  I  D  G  D  F

                1370                1390
          .         .         .         .         .
TTGAATTCGAAACCTATTACTCGCTCTCTTGCCACAACTGCCCGGACGTG
  E  F  E  T  Y  Y  S  L  S  C  H  N  C  P  D  V

      1410                1430                1450
          .         .         .         .         .
GTGCAGGCGCTGAACCTGATGAGCGTACTGAACCCGCGCATCAAGCACAC
V  Q  A  L  N  L  M  S  V  L  N  P  R  I  K  H  T

                1470
          .         .         .
TGCAATTGACGGCGGCACCTTCCAGAACGAAAT
 A  I  D  G  G  T  F  Q  N  E
```

TABLE 3.1

AMINO ACID COMPOSITION OF SCRP-23[a]

Amino Acid	Number of Residues per Subunit	Mol %
D	14	7.487
N	5	2.674
T	24	7.487
S	9	4.189
E	15	8.021
Q	5	2.674
P	8	4.287
G	13	6.952
A	19	10.160
C	2	1.070
V	12	6.417
M	3#	1.604
I	13	6.952
L	11	5.882
Y	5	2.674
F	11	5.882
K	14	7.487
H	5	2.674
R	6	3.209
W	3	1.604

MW = 20,847
\# includes initial M residue which is cleaved from the protein *in vivo*.
a. inferred from DNA sequence.

TABLE 3.2 ***scrp23* CODON USAGE**

First Letter	Second Letter U	C	A	G
C	UUU Phe 0	UCC Ser 5	UAU Tyr 1	UGU Cys 0
	UUC Phe 1	UCC Ser 1	**UAC** Tyr 4	UGC Cys 2
	UUA Leu 2	UCA Ser 0	UAA Ochre 1	UGA Opal 0
	UUG Leu 9	UCG Ser 0	UAG Amber 0	UGG Trp 3
C	CUU Leu 0	CCU Pro 1	CAU His 0	**CGU** Arg 5
	CUC Leu 0	CCC Pro 0	CAC His 5	**CGC** Arg 1
	CUA Leu 0	CCA Pro 1	CAA Gln 0	CGA Arg 0
	CUG Leu 10	**CCG** Pro 6	**CAG** Gln 5	CGG Arg 0
A	AUU Ile 3	ACU Thr 4	AAU Asn 0	AGU Ser 0
	AUC Ile 9	**ACC** Thr 10	AAC Asn 5	AGC Ser 3
	AUA Ile 0	ACA Thr 0	AAA Lys 13	AGA Arg 0
	AUG Met 3	ACG Thr 0	AAG Lys 0	AGG Arg 0
G	**GUU** Val 6	**GCU** Ala 8	GAU Asp 3	**GGU** Gly 5
	GUC Val 1	GCC Ala 1	GAC Asp 11	GGC Gly 7
	GUA Val 5	**GCA** Ala 7	**GAA** Glu 15	GGA Gly 0
	GUG Val 0	**GCG** Ala 3	GAG Glu 0	GGG Gly 0

Fop = 0.91
Bold type denotes optimal codons defined by Ikemura (1985).

3.5.4 Nucleotide Sequence Comparisons

Comparison of the *scrp23* nucleotide sequence with the GenBank and EMBL databases revealed strong similarity (86.9% identity over 1462 bp) with the *ahp* locus of *S.typhimurium* (Tartaglia *et al* 1990). This encodes the C22 and F52a subunits of alkyl hydroperoxide reductase, an enzyme which detoxifies the cell of organic hydroperoxides generated during oxidative stress by peroxide. A comparative analysis of the *ahp* loci of *E. coli* and *S. typhimurium* is summarised in figure 3.10.

The *scrp23* gene was found to be 91.8% identical to the *ahpC* gene which encodes the C22 protein in *S. typhimurium*, whilst the downstream reading frame is 88.5% identical to the first 486 nt of the *ahpF* gene which encodes the F52a protein and is situated downstream of *ahpC*. Both the common gene organization and the high nucleotide sequence identity between the two loci suggest that *scrp23* and the downstream reading frame encode the C22 and F52a subunits, respectively, of Ahp in *E. coli*. *scrp23* and the downstream reading frame will be referred to hereafter as the *ahpC* and *ahpF* genes, respectively, of *E. coli*.

Inspection of the *E. coli* and *S. typhimurium ahpC* sequences revealed that the *E. coli ahpC* gene has two insertions within its coding region, of C at position 552 and G at position 718, corresponding to positions

523 and 688 respectively in the *S. typhimurium ahpC* sequence. Both insertions were verified by several sequencing runs on both DNA strands (data not shown). As a result the predicted polypeptides encoded by the reading frames determined for *E. coli* and *S. typhimurium* share significant identity (91.4%) only within the N-terminal 119 residues, which are encoded by DNA sequences preceding the "frameshift" insertion at position 552 of *E. coli ahpC*.

The discrepancy between the putative reading frames of the two sequences could most easily be explained by errors within either of the determined *ahpC* sequences. To investigate this further nucleotides C552 and G718 were inserted into or deleted from the *S. typhimurium* or *E. coli ahpC* sequences, respectively, and then the codon usage and codon capacity of the alternate reading frames were examined. For convenience sake, the reading frame predicted by the sequence determined for *ahpC* in *S. typhimurium* is termed frame I whereas the reading frame apparent in the *E. coli* sequence is termed frame II.

The codon usage of frames I and II was examined for *ahpC* in *S. typhimurium* using the codon frequency programme. The results are shown in Table 3.3.

Frame I gave a relatively low Fop value of 0.66, whereas the Fop for frame II was found to be 0.85. This is

more characteristic of an efficiently expressed protein (such as is encoded by *ahpC*), and is similar to the 0.91 value obtained for *ahpC* in *E. coli* (see Section 3.5.3).

As summarised in Table 3.4 the predicted C22 polypeptide encoded by frame I in *E. coli* is truncated by 24aa relative to that encoded by frame I in *S. typhimurium*, due to a second position substitution in codon 160 changing it from a TTG (Leu) codon in *S. typhimurium* to a TAG (stop) codon in the *E. coli* sequence. The resulting 159aa polypeptide, of predicted MW 17652, is much smaller than suggested by its mobility on SDS-PAGE. In contrast the C22 polypeptides encoded by frame II in both organisms are predicted to be the same length (187aa including f-Met). Frame II is therefore considered to be the most likely to encode the C22 protein in both organisms.

Figure 3.10

Summary of *ahp* sequence comparisons. Open boxes represent the coding regions of the respective genes which are identified above. Arrows indicate the direction of transcription of each gene.

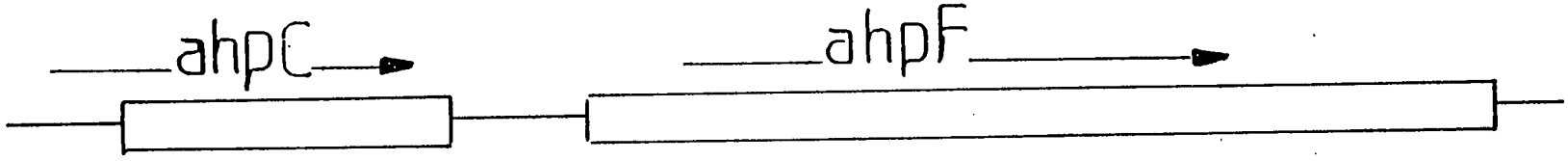
A. Length in base pairs of the genes (excluding stop codons) and of the intergenic regions.

B. % identity of the DNA sequence.

C. Extent of the known sequence of the *E. coli ahpF* gene.

Symbols: Ec denotes *E. coli*; St denotes *S. typhimurium*.

FIGURE 3.10



bpEc.[a]	561	241	486[c]
bpSt	561	245	1536
%DNA[b]	91·8	79·6	88·5

TABLE 3.3 **CODON USAGE OF *ahpC* in *S. typhimurium* #**

First Letter	Second Letter															
	U		I	II	C		I	II	A		I	II	G		I	II
U	UUU	Phe	2	3	UCU	Ser	2	2	UAU	Tyr	3	2	UGU	Cys	0	0
	UUC	Phe	8	8	UCC	Ser	4	3	**UAC**	Tyr	2	3	UGC	Cys	2	2
	UUA	Leu	3	2	UCA	Ser	2	1	UAA	Ochre	1	1	UGA	Opal	0	0
	UUG	Leu	4	0	UCG	Ser	2	0	UAG	Amber	0	0	UGG	Trp	3	3
C	CUU	Leu	1	0	CCU	Pro	3	1	CAU	His	1	1	**CGU**	Arg	3	5
	CUC	Leu	1	0	CCC	Pro	0	0	CAC	His	4	4	**CGC**	Arg	2	1
	CUA	Leu	0	0	CCA	Pro	0	2	CAA	Gln	0	0	CGA	Arg	3	0
	CUG	Leu	7	9	**CCG**	Pro	5	5	**CAG**	Gln	4	5	CGG	Arg	2	0
A	AUU	Ile	2	3	ACU	Thr	4	8	AAU	Asn	1	5	AGU	Ser	1	3
	AUC	Ile	5	8	**ACC**	Thr	9	5	AAC	Asn	6	0	AGC	Ser	3	0
	AUA	Ile	0	0	ACA	Thr	0	0	AAA	Lys	9	13	AGA	Arg	1	0
	AUG	Met	3	3	ACG	Thr	3	1	AAG	Lys	7	1	AGG	Arg	2	0
G	GUU	Val	3	7	**GCU**	Ala	0	3	GAU	Asp	4	4	**GGU**	Gly	4	6
	GUC	Val	3	4	GCC	Ala	5	4	GAC	Asp	7	10	GGC	Gly	5	7
	GUA	Val	4	7	**GCA**	Ala	3	5	**GAA**	Glu	9	14	GGA	Gly	1	0
	GUG	Val	2	0	**GCG**	Ala	4	7	GAG	Glu	1	1	GGG	Gly	0	0

Fop Frame I = 0.66
Fop Frame II = 0.85

TABLE 3.4 SUMMARY OF THE CODING CAPACITY OF ALTERNATIVE *ahpC* READING FRAMES

	FRAME I		FRAME II	
E. coli DNA[a]	477		561	
S. typhimurium	549		561	
E. coli aa[b]	159	(17652)	187	(20,846)
S. typhimurium	183	(20699)	187	(20,747)
% aa[c]	91.8		97.9	

a) Length in bp of *ahpC* coding sequence excluding stop codons.
b) Length of *ahpC* gene products including initial methionine.
c) % identity between *S. typhimurium* and *E. coli ahpC* gene products. Figures in brackets denote the MW of gene products in Daltons.

3.6 Transcriptional Analysis of the *ahp* Genes

Transcription of the *ahpCF* genes of *E. coli* and *S. typhimurium* is elevated during peroxide stress through the action of the OxyR transcriptional regulator protein (Tartaglia *et al* 1989). This actives transcription by RNA polymerase *in vitro* by binding to a site ~45bp in length, upstream of and overlapping the -35 elements of the *S. typhimurium ahp* promotor and the *katG* and *oxyS* promoters of *E. coli* (Tartaglia *et al* 1989, Storz *et al* 1990, Tao *et al* 1991). OxyR also negatively regulates transcription of its own gene (*oxyR*) and that of the phage Mu *mom* gene. In this role, both oxidised and reduced forms of OxyR appeared to be effective (Christman *et al* 1989, Tao *et al* 1991).

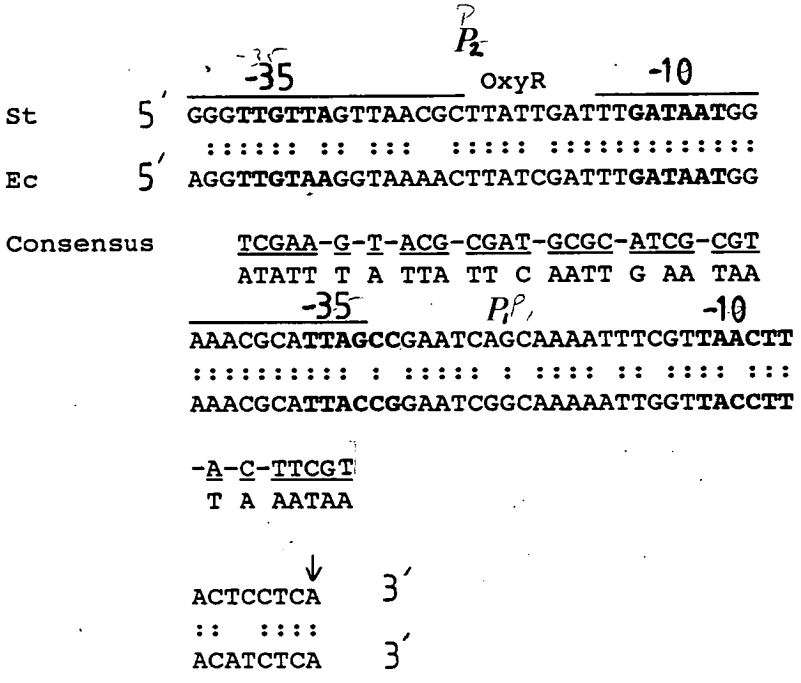
3.6.1 Nucleotide Sequence of the *ahp* Promoter Region

Examination of the nucleotide sequence of the DNA upstream of the *E. coli ahpC* gene revealed two possible sigma 70 dependent promoter sequences. The more proximal (downstream) of these $ahpP_1$ comprising a -10 element (TACCTT) centred around nt 161 and a -35 element (TTACCG) near nt 138, is similar both in sequence and location to the -10 and -35 elements of the OxyR dependent *ahp* promotor of *S. typhimurium* (figure 3.11). The putative distal promoter $ahpP_2$, comprising -10 (GATAAT) and -35 (TTGTAA) elements centred near nt 122 and 101 respectively, is also present within the

Figure 3.11

Alignment of the promoter regions of the *ahpC* genes of *S. typhimurium* (S.t.) and *E. coli* (E.c.). : marks identical nucleotides. The -35 and -10 elements of the known *ahpC* promoter in *S. typhimurium* and putative promoters upstream of *ahpC* in both organisms are shown in bold type; ↓ indicates the start site of transcription from P *ahp* in *S. typhimurium*. The extent of the OxyR footprint at the *S. typhimurium* promoter is indicated above the alignment (Tartaglia *et al* 1989) and the degenerate consensus OxyR binding sequence proposed by Tartaglia *et al* (1992) is shown below the alignment.

Figure 3.11


P2
-35
OxyR
-10
St 5′ GGGTTGTTAGTTAACGCTTATTGATTTGATAATGG
Ec 5′ AGGTTGTAAGGTAAAACTTATCGATTTGATAATGG
Consensus
TCGAA-G-T-ACG-CGAT-GCGC-ATCG-CGT
ATATT T A TTA TT C AATT G AA TAA
-35
P1
-10
AAACGCATTAGCCGAATCAGCAAAATTTCGTTAACTT
AAACGCATTACCGGAATCGGCAAAAATTGGTTACCTT
-A-C-TTCGT
T A AATAA
ACTCCTCA 3′
ACATCTCA 3′

S. typhimurium ahp promoter region (within the OxyR binding site). However, P2 was not utilised in *S. typhimurium* under the conditions tested (Tartaglia *et al* 1989). There is, in fact a high degree of identity between the DNA sequence upstream of the *E. coli* ahpC gene and that of the *ahp* promoter region of *S. typhimurium*, with only seven nucleotide substitutions in a 46bp region which corresponds to the known OxyR binding site in *S. typhimurium*. Furthermore, only one of the seven nucleotide substitutions within this region affects a position of degenerate homology between OxyR targets as defined by Tartaglia *et al* 1992).

3.6.2. Primer Extension Analysis of *ahp mRNA*

Either or both of the putative promoters described in Section 3.6.1 may allow transcription of the *E. coli* *ahp*CF genes. In order to determine whether either of the two putative promoters is active *in vivo*, the 5' termini of *ahpC* transcripts from aerobically grown or H_2O_2-stressed *E. coli* cells were determined by primer extension analysis of total cellular RNA.

RNA was prepared from *E. coli* strain W3110 or *S. typhimurium* strain LT2 (grown normally or under H_2O_2 stress) as described in Section 2.5.7, and subjected to primer extension analysis using [^{32}P] labelled oligonucleotide 482X [5'-GGTAGAAGAAGAAGACGCTCCAGCGGCCTTCGG-3'] described in Section 2.8.7. 482X is complementary to nt 312 through 280 of the coding strand of the *ahpC* gene of *E. coli* and nt 294 through 271 of the same gene in *S. typhimurium*. *S. typhimurium* RNA thus served as a positive control, since its *ahpC* transcription start site had previously been determined by Tartaglia *et al* (1989).

Two major transcripts were identified as initiating within the promoter region of *ahpC* during normal aerobic growth of *E. coli* (figure 3.12, track 1). Reverse transcript I represents transcription initiation at nucleotide A117, 8bp downstream of the -10 box of putative promoter $ahpP_1$ in the *ahpC* sequence, and

corresponds to the single reverse transcript present in track 3, which is the cDNA product expected from mRNA initiated at the known *ahp* promoter in *S. typhimurium* (Tartaglia *et al* 1989). Reverse transcript II (track 1) is the cDNA product of mRNA initiated at G131, 7bp downstream of the -10 box of putative promoter $ahpP_2$ in the *E. coli ahp* sequence. The corresponding transcript was not synthesised from *S. typhimurium* RNA, indicating that P_2 is not active during normal growth in *S. typhimurium*.

ahp transcripts from *S. typhimurium*, and transcripts initiated from $ahpP_1$ in *E. coli*, were elevated after 15 minutes treatment with H_2O_2 (65μM), as indicated by tracks 2 (*E. coli*) and 4 (*S. typhimurium*) in figure 3.12, while transcriptional initiation from P_2 in *E. coli* was repressed (as indicated by the absence of transcript II in track 2). By analogy with the situation in *S. typhimurium*, the elevated levels of transcript I in *E. coli* cells after treatment with H_2O_2 suggest that $ahpP_1$ is an OxyR activated promoter in *E. coli* and that $ahpP_2$ is sterically repressed by OxyR during peroxide-stress.

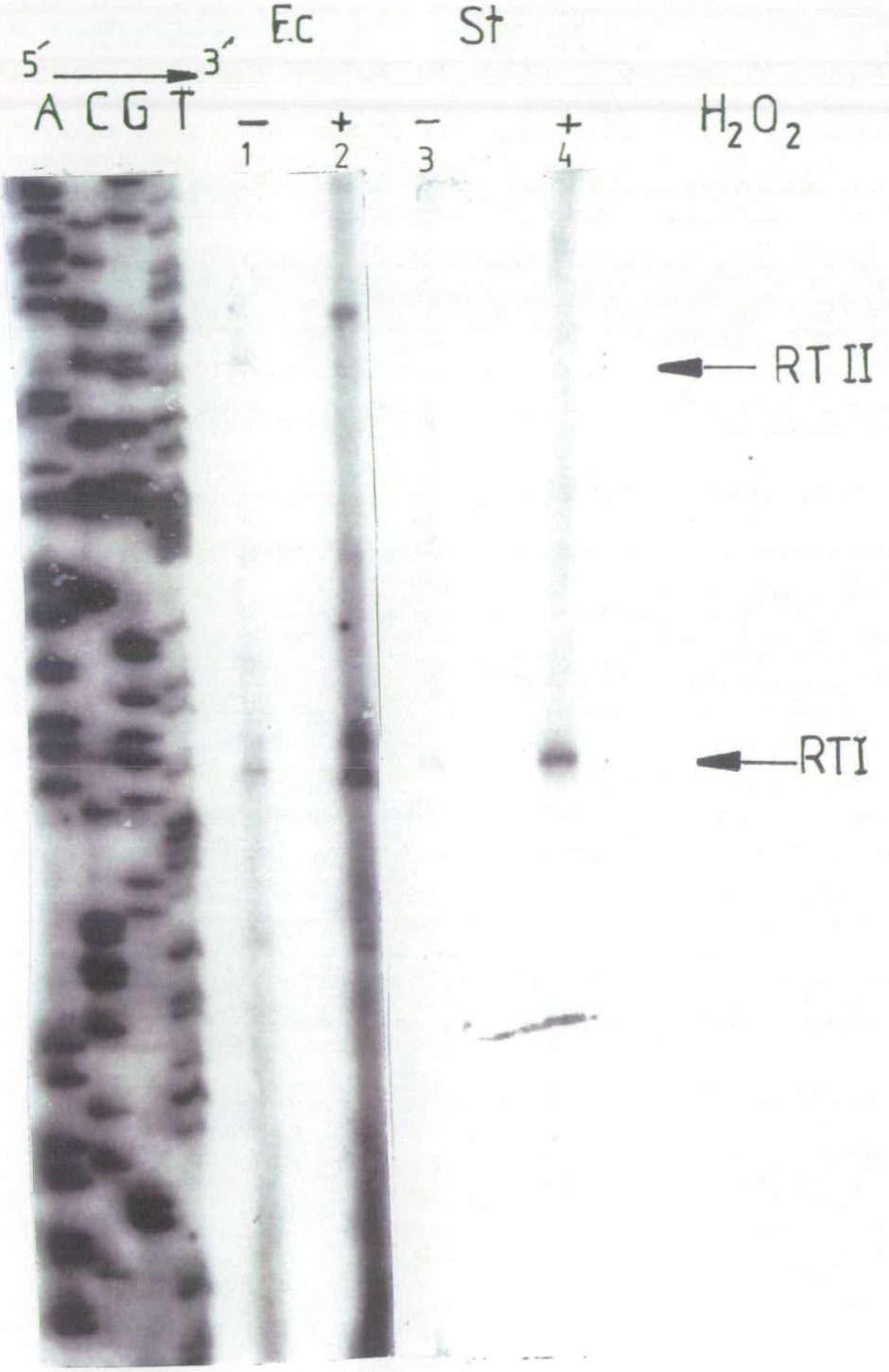
To confirm that $ahpP_1$ is an OxyR-activated promoter, a plasmid -pDSA24-(see Section 3.8.1), which carries the *E. coli ahpC* gene plus upstream promoter region DNA, was sent to the laboratory of Professor Akira Ishihama at

N.I.G., Japan, who tested the ability of OxyR to activate the promoter by transcription assays *in vitro*. The results (data not shown) confirmed that $ahpP_1$ is activated by OxyR and that $ahpP_2$ is repressed by active OxyR protein. (The ability of reduced (inactive) OxyR to repress $ahpP_2$ activity *in vitro* was not tested). Secondly, they showed that $ahpP_2$ is the predominant promoter in the absence of OxyR. Note that transcripts initiated at *E. coli* $ahpP_1$ were not detected in the absence of OxyR *in vitro* (Tao *et al* 1993).

Figure 3.12

Primer extension analysis of *ahpC* transcripts. 30µg of RNA, extracted from *E. coli* strain W3110 (Ec) or *S. typhimurium* strain LT2 (St) which had been grown in the presence or absence of H_2O_2 (65µM), as indicated by + and - above the autoradiograph, were hybridised with the oligonucleotide primer 482X and the primer extended with reverse transcriptase. The hybridisation temperature of 57°C had previously been determined empirically. The resulting cDNA fragments were analysed on a 6% polyacrylamide gel. A DNA sequence ladder primed with the same oligonucleotide, using M13 mp19-22 ssDNA as template, served as marker. RTI and RTII point to the cDNA bands which correspond to the start sites of *ahpC* transcription.

FIGURE 3.12



3.7 A.A. Sequence Comparisons

3.7.1 The C22 Protein

Alkyl hydroperoxide reductase activity requires both C22 and F52a components of Ahp in *S typhimurium* and *E. coli* (Jacobson *et al* 1989). Transfer of electrons from NAD(P)H to the alkyl hydroperoxide substrate is thought to be mediated by redox-active disulphides on both subunits. The C22 protein is thought to be the site of substrate peroxide reduction, which is coupled to NAD(P)H oxidation on the F52a subunit. However, little is known about the exact function of the C22 protein, which may also play some role in initial substrate binding and targetting (Tartaglia *et al* 1990). Previous Database searches had found no significant identities between the C22 protein and known protein sequences. The following sections summarise information gained from sequence comparisons and analysis of the amino acid sequence of the C22 protein.

The amino acid composition of the C22 protein sequence, determined in Section 3.5.3, revealed that two Cys residues are present. These are Cys 47 and Cys 166, and may represent the redox-active disulphide predicted by the chemical evidence of Jacobson *et al* (1989).

Tartaglia *et al* (1990) found no significant similarity between the *S. typhimurium* C22 amino acid sequence and

the sequences of previously reported genes. However, the published C22 protein sequence may have been partly inaccurate (due to frameshift errors in the nucleotide sequence of the *S. typhimurium ahpC* gene {see Section 3.5.4}). Therefore more recent versions of the protein Databases were searched for similarities to the predicted aa sequence of the *E. coli* C22 protein.

The C22 aa sequence was compared with the SWISS-PROT NBRF-PIR and translated GenBank and EMBL and Databases using the Prosearch programme run on SEQNET (see Section 2.11). As Table 3.5 shows, the C22 aa sequence showed significant identity with a number of sequences (13 in total) in the Databases, ranging between 97.9% identity with the revised *S. typhimurium* C22 protein sequence and 20.5% identity with *E. coli* Bacterioferritin co-migratory protein. Further pairwise alignments (Table 3.6) clearly demonstrated that each of these aa sequences shared significant identity with the other 13 protein sequences.

Optimal alignment of all of the matching sequences using the GCG Pileup Programme is represented in figure 3.13. All 13 proteins in the alignment share regions of similarity and have been divided into two groups based on overall sequence identity and the occurrence of these shared regions within each individual sequence.

TABLE 3.5 DATABASE SEQUENCES SIMILAR TO THE *E. coli* C22 PROTEIN

Sequence	Length (Residues)	MW[a]	Genbank Accession Number	Species + Protein	% Identity[b]	Source and Reference
1 eco_ahpc	187	20,846	D13187	*Escherichia coli* Alkylhydroperoxide reductase C22 sub-unit	100	This work
2 ahpc_salty	187[c]	20,747	J05478	*Salmonella typhimurium* Alkylhydroperoxide reductase C22 sub-unit	97.9	GenBank Tartaglia *et al* (1990)
3 YNDH_BACSP	168[d]	18,479	D01111	*Bacillus alcalophilus* orf 5' to NADH dehydrogenase gene (Translation Product)	68.5	SWISS-PROT Xu *et al* (1991)
4 AXYNOX	98[d]	10,560	D13563	*Amphibacillus xylanus* orf 5' to NADH oxidase gene (Translation Product)	65.3	GenBank Niimura (unpublished)
5 MER5_MOUSE	257	28,127	M28723	Mouse Mer 5 protein Housekeeping type	36.9	SWISS-PROT Yamamoto *et al* (1989)
6 R20K_CLOPA	178	20,036	M60116	*Clostridium pasteurianum* 20 kDa protein product of rubredoxin operon orf 3	34.3	SWISS-PROT Mathieu *et al* (1992)
7 CR29_ENTHI	273	25,333	M35635	*Entamoeba hystolitica* 29 kDa cysteine rich surface antigen	30.5	GenBank Torian *et al* (1990)
8 26KD_HELPY	198[e]	22,259	M55507	*Helicobacter pylori* 26 kDa antigen	29.4	SWISS-PROT O'Toole *et al* (1991)

TABLE 3.5 DATABASE SEQUENCES SIMILAR TO THE *E. coli* C22 PROTEIN (Continued)

Sequence	Length (Residues)	MW[a]	Accession Number	Species + Protein	% Identity[b]	Source and Reference
9 S85445	195	21,655	M74232	*Mycobacterium avium* Avi-3 antigen	34.8	GenBank Yamaguchi *et al* (1992)
10 MSGB38COS	195	21,469	L01095	*Mycobacterium leprae* genomic DNA sequence "cosmid b38" (Translation Product)	31.6	GenBank Smith (Unpublished)
11 MBF_SOD	174[d,f]	19,834	D00614	*Methanobacterium thermoautotrophicum*; orf 5' to superoxide dismutase gene (Translation Product)	24.1	GenBank Takao *et al* (1990)
12 BMU_EMSP	202	22,380	X63202	*Bromus secalinus* polypeptide product of embryospecific mRNA	22.5	NBRF-PIR Goldmark *et al* (1992)
13 BCP_ECOLI	156	17,634	M37689	*Escherichia coli* bacterioferritin co-migratory protein	20.5	SWISS-PROT Andrews *et al* (1991)

a. MW calculated using GCG PEPTIDESORT Programme (no enzymes).
b. % identity with eco ahp.
c. Sequence has been modified to include frame-shift corrections proposed in Section 3.5.4 of this thesis.
d. N-terminally truncated aa sequence.
e. C-terminally truncated aa sequence.
f. Sequence contains frame-shift corrections proposed by S. J. McQuay (PhD thesis, University of Edinburgh, in preparation) which improve and extend sequence identity between MBF_SOD and ecoahpc.

This table was compiled using data collected by the author in collaboration with S. J. McQuay of the Biocomputing Research Unit of this Institute. All or part of these data may appear in the latter's PhD thesis (University of Edinburgh, in preparation).

TABLE 3.6

PAIRWISE IDENTITIES OF SEQUENCES RELATED TO THE C22 PROTEIN[#]

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		97.9	68.5	65.3	36.9	34.3	30.5	29.4	34.9	31.6	24.1	22.5	20.5
2			69.1	64.3	36.4	34.3	30.0	29.4	34.2	32.1	24.7	22.5	21.2
3				83.7	39.3	31.6	35.7	32.7	32.1	31.6	23.8	21.4	19.2
4					38.8	26.5	38.8	35.7	28.6	27.6	20.4	16.3	13.3
5						51.7	41.2	45.5	35.4	33.9	23.6	19.8	24.4
6							44.4	39.9	40.6	39.9	23.6	22.5	22.4
7								39.4	29.7	28.2	24.7	20.8	25.0
8									24.9	33.3	21.3	24.2	25.0
9										85.1	22.4	22.6	13.5
10											21.3	22.6	18.6
11												32.2	11.5
12													16.7
13													

[#] Data was collected using GCG-distances. Identity is shown as a percentage of the length of the shorter of the two protein sequences, in each pairwise comparison (GCG-distances Denominator: "Length of shorter sequence without gap"). Nos. 1-13 correspond to sequences in Table 3.5

Data was compiled in collaboration with S. J. McQuay and may appear in this or other form(s) in her PhD thesis (University of Edinburgh in preparation).

The core group of 10 protein sequences (bracketted in figure 3.13) share 4 regions of strong sequence similarity (including a high number of identical residues). The most striking of these, region 1, is 14aa in length and corresponds to aa 38 through 50 of the C22 protein sequence. Included within region 1 is a very striking prevalence of aromatic aa, and an invariant Cys residue (Cys 47 in C22) which may form part of the predicted redox-active disulphide of the C22 protein. A second invariant Cys (Cys 166 in the C22 protein) is located within region 4 of the core group. Three members of the core group, R20K CLOPA, S85445 and MSGB38COS show only limited similarity to consensus in region 4, but are considered to be core group members based on the occurrence of both invariant Cys residues.

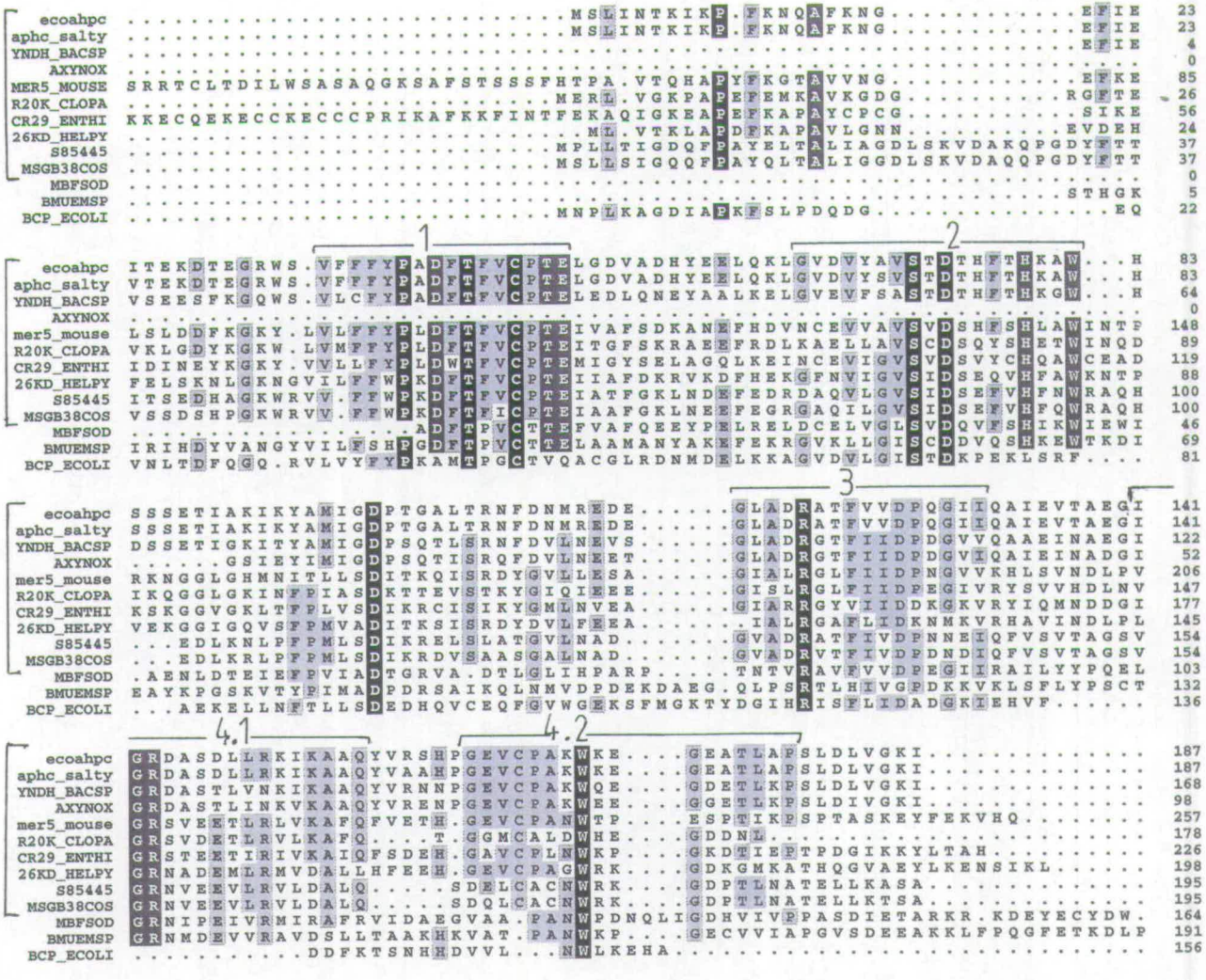
The non core group proteins generally show lower pairwise similarity to the core group proteins than do the latter's constituent members. While an invariant Cys residue corresponding to Cys 47 of the C22 protein is present in all non core group proteins, none of them have a Cys residue within sub-region 4.2, which is also poorly conserved in other respects, and barely recognisable in BCP_*E. coli*. Although the non core group proteins show poor overall sequence similarity with the core group there is significant clustering of invariant amino acid residues within regions 1, 2, 3 and sub-region 4.1 (with the exception of BCP_*E. coli*, from

which most of region 4 is missing). They are therefore retained within the sequence alignment.

It should be noted that N—terminal portions of YNDH_BACSP, AXYNOX and MBFSOD have not yet been reported. This is scarcely critical in the case of YNDH_BACSP as the available sequence includes both regions 1 and 4, each containing an invariant Cys residue. MBFSOD is consigned to the non core group despite the absence of part of the region 1 sequence information, as it shows relatively poor similarity with region 4 and does not include a Cys comparable to Cys166 of the C22 aa sequence. Regions 1 and 2 are missing from the available AXYNOX sequence. However, its strong similarity to core proteins in region 4, and the presence of the invariant Cys residue, tentatively support its inclusion within the core group.

Figures 3.13

Multiple alignment of C22 related protein sequences. The positions of the aa residues at the right end of each row within the corresponding protein sequences are marked to the right of each row. Four regions of strong similarity are indicated (1 through 4). Region 4 is divided into subregions 4.1 and 4.2. Invariant residues are identified by black backgrounds. Residues which are invariant in all proteins except BCP_E.COLI are identified by dark grey backgrounds. Light shading identifies aa residues which are present in at least 7 proteins in the aligment. Insertions made during alignment optimisation are marked by a dot. Proteins in the aligment are identified fully in table 3.5. This figure was kindly produced by S. J. McQuay of the Biocomputing Research Unit of this Institute.


1
2
3
4.1
4.2
ecoahpc
aphc_salty
YNDH_BACSP
AXYNOX
MER5_MOUSE
R20K_CLOPA
CR29_ENTHI
26KD_HELPY
S85445
MSGB38COS
MBFSOD
BMUEMSP
BCP_ECOLI

3.7.2 The F52a Protein

The F52a protein is thought to be a member of a class of disulphide oxidoreductases related to thioredoxin reductase. The C-terminal 206 aa of F52a from *S. typhimurium* shares 35.4% identity with *E. coli* thioredoxin reductase, including a pair of Cys residues known to form a redox-active disulphide in thioredoxin reductase (Tartaglia *et al* 1990). Both proteins also share 3 short regions of local sequence similarity with a second class of disulphide oxidoreductases, including glutathione reductase and dihydrolipoamide dehydrogenases from different organisms, which are known to represent an NAD binding site and two FAD binding sequences (Russell and Model 1988, Tartaglia *et al* 1990). Tartaglia *et al* (1990) found no significant similarity between the N-terminal aa of the F52a protein and previously reported protein sequences. Therefore more recent versions of the protein Databases were searched for sequences similar to the N-terminal "domain" of the F52a protein.

Only the first 162 codons of the *E. coli ahpF* DNA sequence were determined in this study, so the whole of the *S. typhimurium* F52a protein sequence was used to search the SWISS-PROT and NBRF-PIR Databases.

As Table 3.7 shows, the F52a protein sequence was found to show significant similarity to four other sequences

in the Database. These are the NADH dehydrogenase of *Bacillus alcalophilus* (Xu *et at* 1991), NADH oxidase from *A xylanus* (Niimura *et al* 1993), a 34 kDa protein encoded by the rubredoxin operon of *C. pasteurianum* (Mathieu *et al* 1992) and *E. coli* thioredoxin reductase (Russell and Model 1988). The similarity between the F52a protein and thioredoxin reductase has previously been discussed by Tartaglia *et al* 1990. Similarly, Mathieu *et al* (1992) reported the similarity of the 34kDa protein of *C. pasteurianum* to the F52a protein and thioredoxin reductase. Xu *et al* (1991) also noted the similarity between the *B. alcalophilus* NADH dehydrogenase and thioredoxin reductase.

Four of these sequences, plus the N-terminal 162 aa of the *E. coli* F52a protein were multiply aligned using the GCG PILEUP programme (figure 3.14) which in addition to detecting the similarity between the C—terminal domains of the F52a protein of *S. typhimurium*, the *B. alcalophilus* NADH dehydrogenase and *E. coli* thioredoxin reductase, revealed that the F52a proteins of *E. coli* and *S. typhimurium* contain an N—terminal domain, not present in thioredoxin reductase, which shows strong sequence similarity with the membrane binding domain of the *Bacillus* NADH dehydrogenase. A similar domain is present at the N-terminus of the *A. xylanus* NADH oxidase which also shows singificant similarity near its C-terminus to thioredoxin reductase

and the F52a protein of *S. typhimurium* (Niimura *et al* 1993).

Figure 3.14

Alignment of F52a related protein sequences. The sequences are identified as follows: ahpF, F52a of *E. coli*; ahpF_Salty, F52a of *S. typhimurium*; dhna_bacsp, NADH dehydrogenase of *B. alcalophilus*; y34K_clopa, orf 1 gene produce of *C. pasteurianum* rubredoxin operon and trxB, thioredoxin reductase of *E. coli*. The sequences were aligned using GCG PILEUP. Residues which are identical in at least three proteins are shaded. The redox active cysteine residues of the thioredoxin reductase protein and their counterparts in the related sequences are indicated below by +. Proposed FAD and NAD(P)H binding domains (previously identified by Tartaglia *et al* 1990 and Niimura *et al* 1993) are boxed and identified below the alignment. Insertions made during optimisation of the alignment are marked with a dot. Short stretches of moderately hydrophobic aa which might aid the association of ahpF, ahpF_salty and dhna_bacsp to the cell membrane are underlined. The final version of this figure was produced by S. J. McQuay of the Biocomputing Research Group of this Institute and may appear in this or another form in her PhD thesis (University of Edinburgh, in preparation).

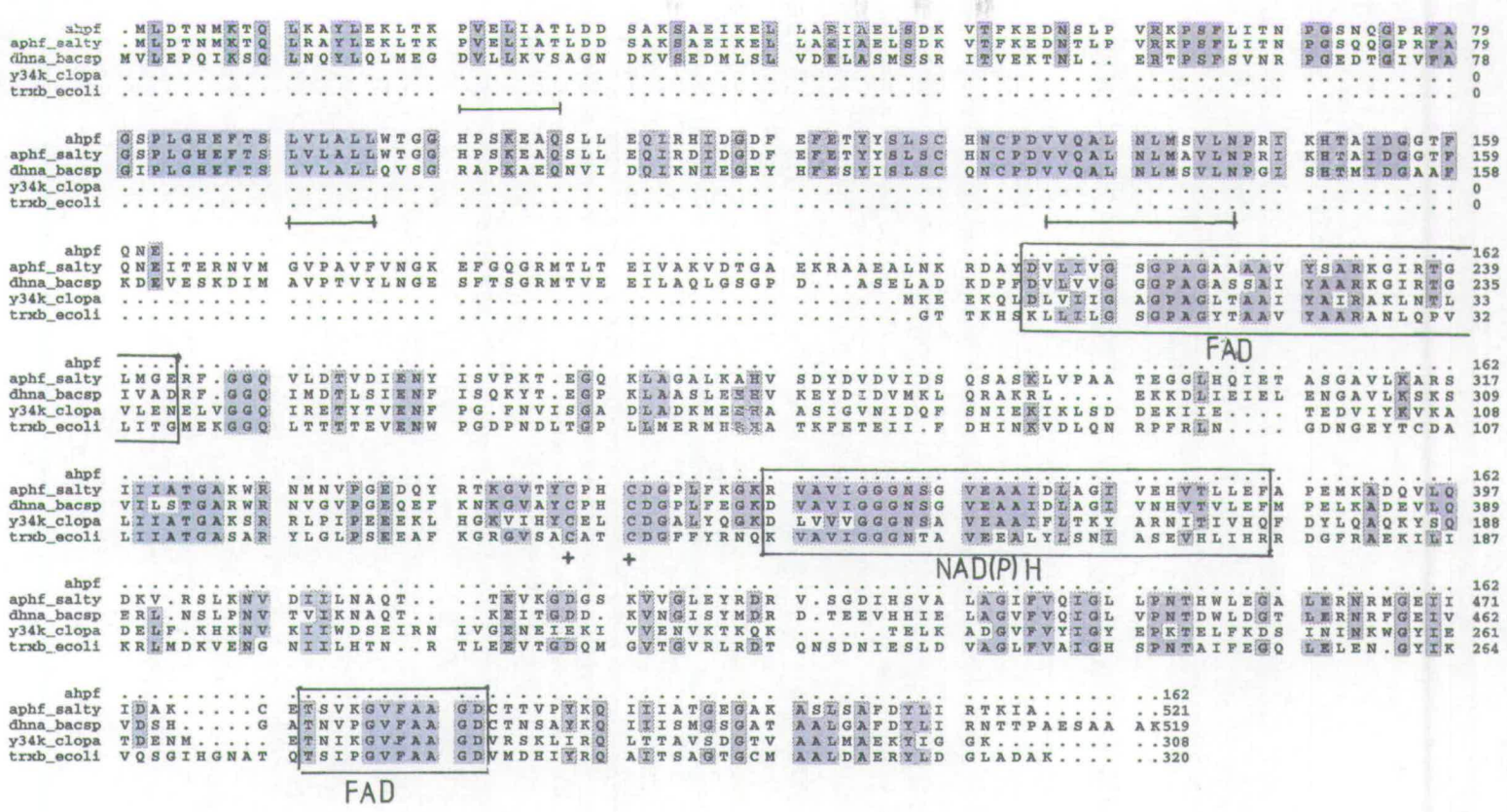


FIGURE 3.14

TABLE 3.7 **Proteins showing Sequence Similarity to the *S. typhimurium* F52a Protein**

Gene Product	Organism	Accession Number	MW[a]	% Identity[b]	Reference
NADH dehydrogenase	*B. alcalophilus*	P26829 (SWISS-PROT)	55830	54.9	Xu *et al* (1990)
NADH oxidase	*A. xylanus*	D13563 (GenBank)	54977	57.2	Niimura *et al* (1993)
Thioredoxin reductase	*E. coli*	P09625 (SWISS-PROT)	34000	34.0	Russell & Model (1988)
*orf*1 gene product	*C. pasteurianum*	P23160 (SWISS-PROT)	32400	31.4	Mathieu *et al* (1992)

a. MW calculated using the GCG PEPTIDESORT programme (no enzymes).
b. Identity to the *S. typhimurium* F52a protein. Identities were calculated using the GCG FastA and TFastA programmes and are shown as a percentage of the length of the shorter sequence.

3.7.3 Quantitation of the Ahp Subunits *In Vivo*

The ~22.5kDa and ~53kDa proteins observed in Section 3.4 are almost certainly respectively the C22 and F52a subunits of Ahp. By inspection of figure 3.6 and 3.7a, the C22 protein appears to be expressed in significant excess over the F52a protein. In order to obtain a more quantitative estimate of the relative levels of expression of the two subunits, autoradiographs obtained in the experiment summarised in Section 3.4.1 were subjected to densitometric scans (Table 3.8). The results, corrected for aa composition (assuming *E. coli* F52a is very similar to the *S. typhimurium* protein throughout) indicate that 9-12 moles of C22 are expressed per 1 mole of F52a protein *in vivo*, in UV irradiated *E. coli* cells infected with lambda 166 or lambda 167. This suggests that Ahp may not function as a simple oligomeric complex, that C22 might conceivably have independent functions, and that the *ahpC* and *ahpF* genes could conceivably be subject to some degree of independent regulation.

It was not possible to accurately quantify the relative levels of the C22 and F52a polypeptides in minicells due to the fact that the C22 and CAT proteins migrated very close to each other during SDS-PAGE.

TABLE 3.8 **Quantitation of the Ahp Subunits by Densitometry**

Track [a]	%F52a[b]	%C22[b]	Met[c]		C22[d]/F52a
			F52a	C22	
2	26.7	73.2	7	2	9.61
3	22.2	77.7			12.25
4	27.7	72.8			9.19

a. Track numbers in Figure 3.6.

b. Density per band is expressed as a percentage of the total in each (vertical) track.

c. No. of methionine residues in the mature protein. The initial methionine residue is cleaved from both proteins *in vivo* (Jacobson *et al* 1989, Ueshima *et al* 1992).

d. Molar ratio of C22 to F52a corrected for Met content. The F52a Met content is assumed to be equivalent to that in the reported aa sequence of the *S. typhimurium* F52a protein (Tartaglia *et al* 1990). The Met content of the C22 protein was taken from Table 3.1 of this thesis.

3.8 Construction of a System for Further Studies on *Ahp* Function

Comparative protein sequence analyses (see Section 3.7.1) placed the C22 polypeptide within a potential family of proteins, from diverse organisms, which share "conserved" Cys residues located within larger regions of sequence similarity. The roles of Cys47 and Cys166 and the significance of the "conserved domains" to Ahp function may therefore be of wide biological significance and obviously require further study. Additionally, the results presented in Section 3.7.3 suggest that the function and expression of the *ahpC* and *ahpF* genes could be partially independent.

A set of *ahp* plasmids was constructed, together with an *ahp* deletion strain of *E. coli*, which should allow further study of the functions of the C22 and F52a proteins *in vivo*.

The plasmid pDSA23 expresses the C22 and F52a subnits of Ahp in minicells *in vivo* and was therefore chosen as the starting material for construction of *ahpC* and *ahpF* expression plasmids, and as a vehicle for chromosomal gene disruption.

3.8.1 Construction of the ahpC plasmid pDSA24

The location of the 2 *Sma*I restriction sites, one within the *ahpCF* intergenic region and one downstream of the *ahpF* gene, (see figure 3.5) facilitated the deletion of the *ahpF* gene by *Sma*I digestion of pDSA23. The large ~6.35kb fragment (which carries the *ahpC* gene) was gel

purified and religated as described in Sections 2.6.4 and 2.6.5 to form pDSA24 (figure 3.15). The structure of pDSA24 was verified by restriction analysis (figure 3.16A).

3.8.2 Construction of the *ahpF* Plasmids pDSA25 and pDSA26

pDSA25 (figure 3.15) was constructed by inserting the 2.3kb *Sma*I fragment from pDSA23 into the *Sma*I site of pBlueScript II (SK+). However, all recombinants recovered carried the *Sma*I fragment in the opposite orientation to that required to place the *ahpF* gene under the control of the *lac* promoter. pDSA26 was therefore constructed by excising the *ahpF* gene (from pDSA25) as a *Hin*dIII-*Bam*HI fragment and inserting it between the same sites in pSKS106 (figure 3.15). The structure of pDSA25 and pDSA26 was verified by restriction analysis (figure 3.16B and C).

As figure 3.15B and D shows, pDSA24 (CmR, *Ori* p15a) and pDSA26 (AmpR, *Ori* pBR322) carry distinct selectable drug resistance markers and compatible replication systems and could therefore be maintained within the same host strain if desired.

3.8.3. Disruption of the *ahp* Locus

Disruption of the *ahpCF* locus was achieved by replacing plasmid-borne copies of the *ahpCF* genes with a kanamycin

resistance cassette *in vitro* and then transducing the mutated locus onto the *E. coli* K12 chromosome via homologous recombination, using lambda 166 as a transducing phage as described by Kulakauskas *et al* (1991).

The entire *ahpC* gene and all but the some 300 downstream bp of the *ahpF* gene were deleted from pDSA23 by *Pst*I digestion. The residual 6.25kb fragment was then "gene-cleaned" and ligated with the ~1.3kb *kan* cassette from pUC4-KISS to form pDSA27 (figure 3.17). Here the *kan* cassette is flanked by at least 0.7kb of DNA from the *ahp* locus on either side, allowing homologous recombination between the plasmid-borne alleles.

The transducing phage was generated by growing plate lysates of lambda 166 on strain TG1 (Rec^+) carrying pDSA27 (as described in Section 2.4.5). NM621 (*recD*) was then transduced to Kan^r by infection with transducing phage at a MOI of 1, as described in Section 2.4.8. 8/8 transductants recovered were Kan^r and Cm^s, and were free from pDSA27 DNA as judged by small scale plasmid preparation (data not shown).

In order to verify that replacement of the chromosomal *ahpCF* genes by the *kan* cassette had occurred, genomic DNA was prepared from 4 transductant strains and subjected to restriction and Southern hybridisation

analyses, using as a probe the 4kb *Hin*dIII-*Kpn*I fragment from pDSA23 labelled with alpha-[^{32}P]dCTP by random priming as described in Section 2.8.1.

As figure 3.18 shows the *Hin*dIII-*Kpn*I probe hybridised to a single 4kb fragment of NM621 DNA digested with *Hin*dIII and *Kpn*I, whereas digestion of the transductant DNAs with the same enzymes generated two probe-positive fragments of approximately 1.8kb and 1.2kb in size. (There is an additional *Hin*dIII site within the *kan* cassette.) Similarly, the failure of the probe to detect the 1.8kb *Pst*I fragment within the transductant DNAs confirms that the *ahpCF* genes have been replaced by the *kan* cassette on the NM621 chromosome. As expected, the sizes of the probe-positive chromosomal *Pst*I fragments flanking the replaced 1.8kb and 0.3kb *Pst*I fragments were unaffected. (The 0.3kb fragment was too small to be detected in figure 3.18.) One of the transductants was selected for further characterisation and named DSA102. These results show that, as expected, the *ahp* genes are inessential to *E. coli* under normal growth conditions.

Peroxide Sensitivity of DSA102

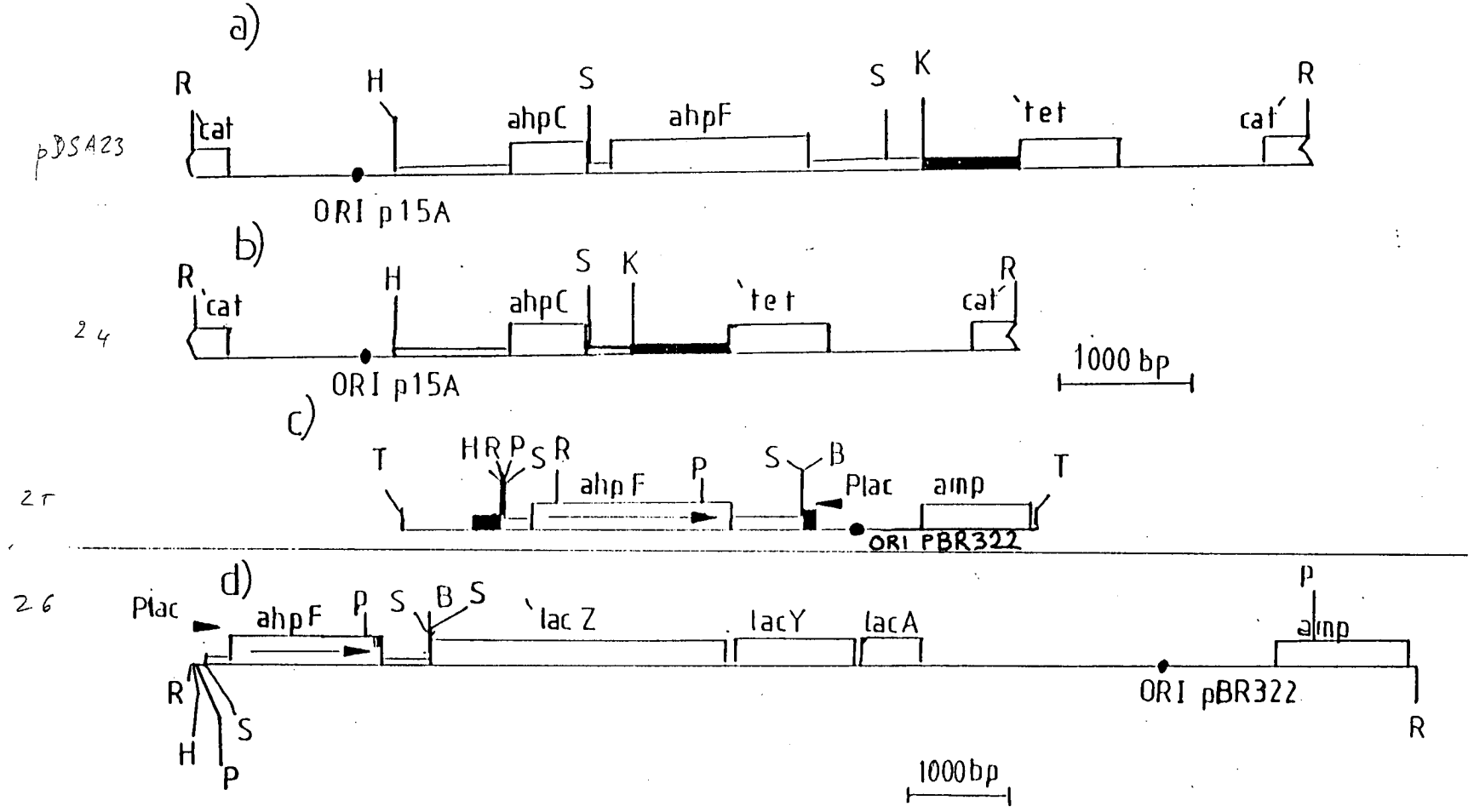
Having verified the replacement of the *ahpCF* genes by the *kan* cassette from pDSA27, in the DSA102 chromosome, it was necessary to check the sensitivity of DSA102 to CHP - a model alkylhydroperoxide which is readily reduced to cumyl alcohol by Ahp *in vitro* (Jacobson *et al* 1989). As described in Section 2.3.4 Ahp activity can be assayed *in vivo* by measuring the sensitivity of *E. coli* strains to CHP in disc inhibition assays. Ahp deficient strains are characteristically hypersensitive when compared to Ahp^+ strains (Storz *et al* 1989, Tartaglia *et al* 1990).

DSA102 and its parent strain, NM621, were grown overnight in LB (containing kanamycin at 50$\mu g ml^{-1}$ where required). 0.1ml aliquots were then assayed for CHP sensitivity on LB agar or glucose minimal agar as described in Section 2.3.4. The results are summarised in Table 3.9. DSA102 was clearly hypersensitive to CHP, as demonstrated by the diameter of the zone of killing which is over 50% greater than that observed for NM621 on **both** LB and minimal agar plates. The greater size of the zone of killing observed on minimal plates for **both** strains was reproducible and could possibly be due to a lower availability of reducing equivalents and/or to more serious effects of oxidative damage to biosynthetic pathways when cells are grown on nutrient-limited media.

Figure 3.15

ahp expression plasmids: a) pDSA23, b) pDSA24, c) pDSA25 and d) pDSA26. Boxes show relevant genes. ▬ represents relevant promoters. Thick black lines represent M13 mp18 DNA(a and b) or *lacZ* DNA(c) respectively. Unfilled thick lines represent ahpC and ahpF flanking sequences. ● denotes plasmid replication origins which are identified for each construct. *amp* encodes beta-lactamase, *cat* encodes chloramphenicol acetyl transferase and *'tet* represents the 5'-truncated tetracycline resistance gene. Only relevant restriction sites are shown for which the following abbreviations were used: B, *Bam*HI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sma*I and T, *Ssp*I. pDSA23 and pDSA24 are shown linearised at the *Eco*RI site (position 1 in the pACYC184 nucleotide sequence); pDSA25 is shown linearised at an *Ssp*I site (position 2850 in pBluScriptII SK+) and pDSA26 is shown linearised at an *Eco*RI site (position 0 in pSKS106).

FIGURE 3.15



191B

Figure 3.16

Restriction analysis of *ahp* expression plasmids.

A) 0.8% agarose gel. Tracks 2 and 4 contain pDSA23 DNA; Tracks 3 and 5 contain pDSA24 DNA digested with restriction enzymes which are indicated above the gel photograph. Track 1 contains 1 kilobase ladder marker DNA (Gibco BRL); the relevant marker fragments are marked and their sizes given (in base bairs) to the left of the gel photograph.

B) 1% agarose gel electrophoresis analysis of pDSA25 digested with restriction enzymes which are indicated above the gel photograph. Track M contains 1 kilobase ladder marker DNA (Gibco BRL); the relevant fragments are marked and their sizes given (in base pairs) to the left of the gel photograph.

C) 0.8% agarose gel. Track 2 shows pDSA23 DNA; Tracks 3 and 5 show pDSA26 DNA and Track 4 shows pSKS106 DNA digested with restriction enzymes which are indicated above the gel photograph. Track 1 shows 1 kilobase ladder marker DNA (Gibco BRL); the relevant fragments are marked and their sizes given to the left of the gel photograph. Note that the ~1.55 kb and ~0.95 kb *Pst*I fragments of pDSA26 are not visible in the gel photograph. Both these fragments were observed when the gel was viewed on a UV transilluminator.

The following abbreviations are used for restriction enzymes:
H, *Hin*dIII; R, *Eco*RI; P, *Pst*I and S, *Sma*I.

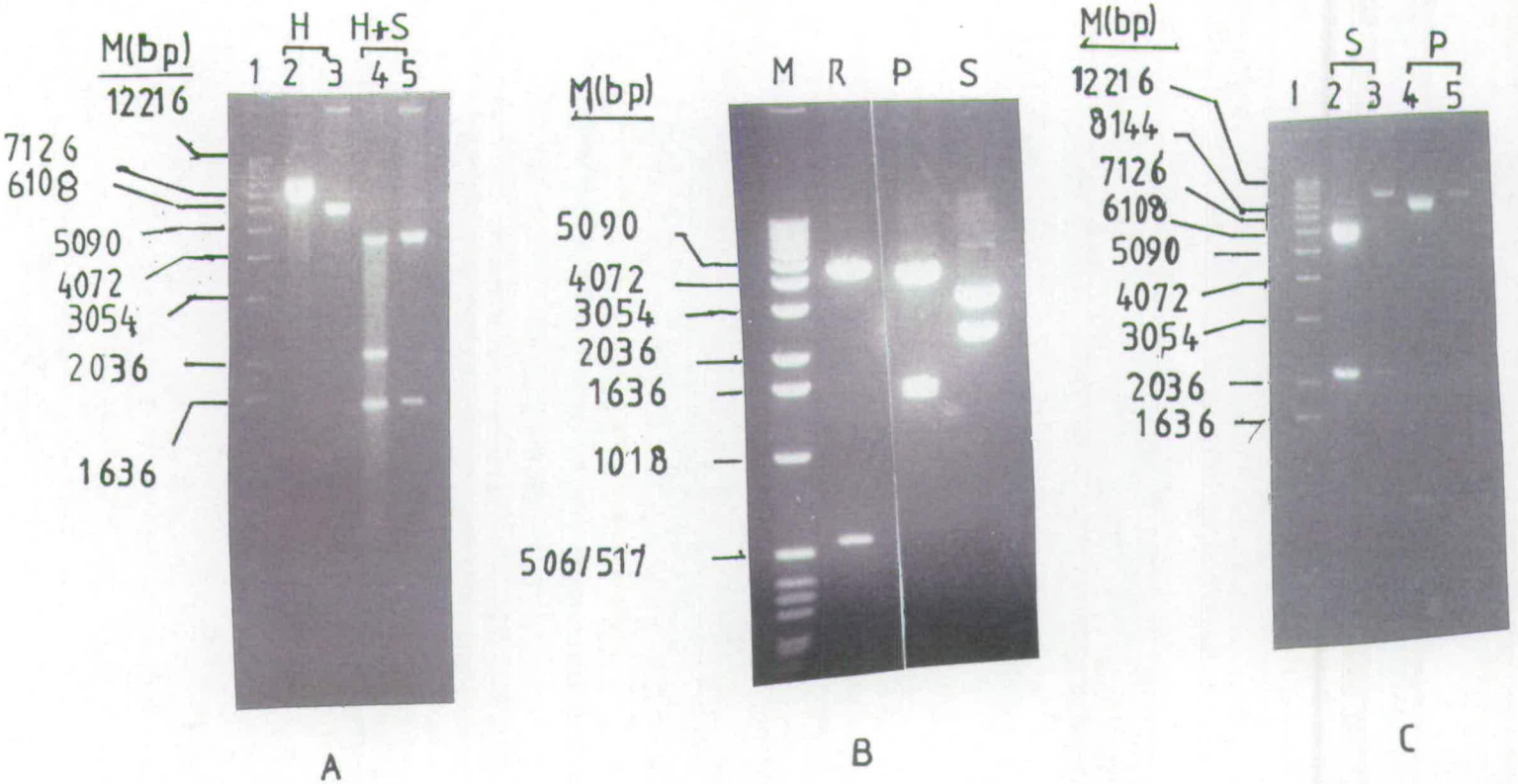

M(bp)
12216
7126
6108
5090
4072
3054
2036
1636
H
H+S
1 2 3 4 5
A
M(bp)
5090
4072
3054
2036
1636
1018
506/517
M R P S
B
M(bp)
12216
8144
7126
6108
5090
4072
3054
2036
1636
S
P
1 2 3 4 5
C

FIGURE 3.16

Figure 3.17

Summary of the construction of pDSA27. Open rectangles represent the relevant genes; hatched boxes represent the 1.3 kilobase kanR cassette derived from pUC4-KISS; thick unfilled lines represent *ahpCF* flanking sequences and thick filled lines represent M13 mp19 DNA. *cat* encodes chloramphenicol acetyl transferase; KanR encodes kanamycin resistance and 'tet represents the 5'-truncated tetracycline resistance gene. Only relevant restriction sites are shown for which the following abbreviations are used: H, *Hin*dIII; K, *Kpn*I; P, *Pst*I and R, *Eco*RI.

FIGURE 3.17

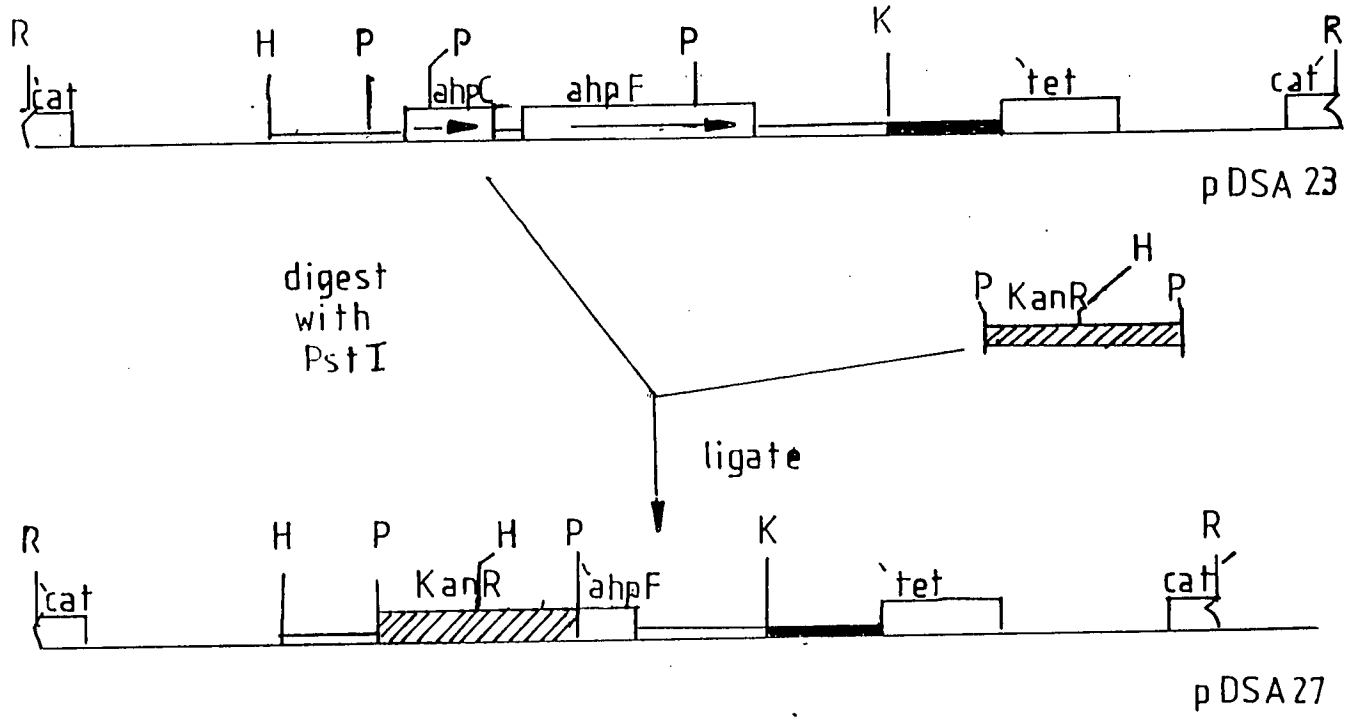

R
H
P
P
P
K
R
cat
ahpC
ahpF
tet
cat
pDSA 23
digest
with
PstI
H
P
KanR
P
ligate
R
H
P
H
P
K
R
cat
KanR
ahpF
tet
cat
pDSA 27

Figure 3.18

Verification of replacement of *ahp* by Kan^r on the E. coli K12 chromosome by Southern hybridisation analysis. Hybridisation and washing were carried out as described in Section 2.8.4. Tracks 2 and 8 show NM621 DNA, Tracks 3, 4, 5, 7, 9, 10, 11 and 12 show transductant DNAs. Track 1 contains 1 kilobase ladder marker DNA (Gibco BRL); the probe-positive 1.636 kilobase marker fragment is derived from pBR322 (all other marker bands marked to the left of the gel photograph are multimers of a 1014 base pair DNA fragment); the sizes of marker fragments are shown in base pairs. The following abbreviations are used for restriction enzymes:
H, *Hin*dIII; K, *Kpn*I and P, *Pst*I.

FIGURE 3.18

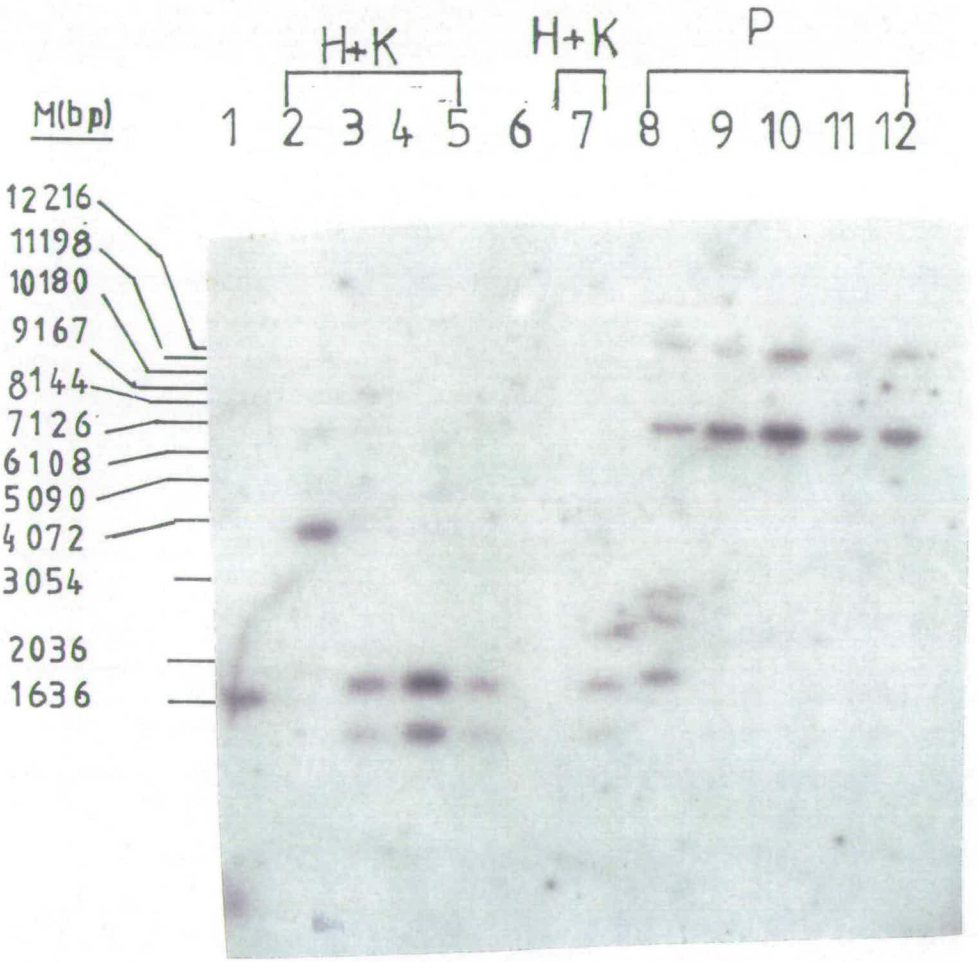

H+K
H+K
P
M(bp)
1 2 3 4 5 6 7 8 9 10 11 12
12216
11198
10180
9167
8144
7126
6108
5090
4072
3054
2036
1636

TABLE 3.9 **SENSITIVITY OF DSA102 TO CUMENE HYDROPEROXIDE (CHP)**

Strain	Kill Zone [a]	
	LB	MM [b]
NM621	16.7	30
DSA102 [c]	26	47

a) Diameter of the zone of killing in millimeters (values stated are the means of three separate observations).

b) Spizizen minimal medium supplemented with glucose, proline, leucine and thiamine at concentrations stated in Section 2.1.3.

c) Kanamycin was added to medium at 50 μg ml^{-1} final concentration for maintenance of the Kan^r phenotype.

In order to verify that pDSA23, pDSA24 and pDSA26 express functional C22 and/or F52a proteins, it was decided to test their ability to complement the CHP sensitive phenotype caused by *ahp* deletion. Unfortunately plasmids carrying ColE1 type replication origins, such as pDSA26, are unstable in *recD* strains. Therefore I first had to move the *ahp:: kan* marker from DSA103 to a new $RecD^+$ background in order to perform complementation tests with pDSA series plasmids.

Construction of DSA103

A Plvir plate lysate was grown on DSA102 as described in Section 2.4.2 and used to transduce *E. coli* K12 strain A19 (*rna*) to Kan^r as described in Section 2.4.3. Kanamycin resistant colonies were selected on LB agar plates containing the drug at a final concentration of $50\mu g ml^{-1}$. In order to check that the *ahp:: kan* marker had been transferred to the correct chromosomal location 50 resistant colonies were replica-patched on to a pair of LB plates (containing kanamycin at $50\mu g ml^{-1}$) and, after growth at 37°C overnight, tested for RNase I activity as described in Section 2.3.5. The short genetic distance between *ahp* and *rna* (~0.2 minutes) means that the two markers should have been cotransduced at high frequency on to the A19 chromosome. Indeed, ~75% (37/50) of Kan^r transductants were also $Rnase^+$ suggesting that the *kan* and *rna* markers are closely linked on the transducant chromosome. Three of the Kan^r $Rnase^+$ transductants were purified and checked for CHP hypersensitivity as described in Section 2.3.4. All 3 transductants were found to be hypersensitive to CHP compared to A19 (data not shown) and one of these was picked for further study and named DSA103.

Suppression of CHP Sensitivity of DSA103

Both the C22 and F52a subunits are required for alkylhydroperoxide reduction *in vitro*. I decided to confirm the requirement for both proteins *in vivo* as a means of testing whether the plasmid pDSA23, pDSA24 and pDSA26 express functional C22 and/or F52a protein. DSA103 was transformed with pDSA23, pDSA24 or pDSA26, or by both of the latter plasmids and tested for CHP sensitivity by disc inhibition assay as described in Section 2.3.4. As shown in Table 3.10 pDSA23, or pDSA24 in combination with pDSA26, were able to suppress the CHP hypersensitivity of DSA103. Neither pDSA24 nor pDSA26 alone was able to do so, indicating that all 3 *ahp* plasmids express functional Ahp proteins and that *in vivo*, as *in vitro*, both the C22 and F52a proteins are required for alkylhydroperoxide reduction. The apparent non-requirement for IPTG for F52a expression from pDSA26 may well reflect titration of LacI repressor by the high copy plasmid. The host DSA103 carries only a wild type chromosomal copy of *lacI*. Indeed, the zones of killing recorded for DSA103 harbouring pDSA26 or pDSA24 and pDSA26 were increased by the presence of IPTG in the medium. This suggests that over-expression of the F52a protein is harmful, perhaps through leading to exhaustion of NADPH which is an important source of reducing equivalents in oxidatively-stressed cells. In relation to this, Storz *et al* (1989) reported that plasmids carrying the *ahpF* gene could not be maintained in an $OxyR^{(cons)}$ strain.

Whilst the results above are consistent with the suggestion that pDSA23, and pDSA24 plus pDSA26 in combination are able to complement

the CHP sensitivity of DSA103 it should be noted that this strain is $RecA^+$. This raises the possibility of marker rescue by recombination of the plasmid borne *ahp* genes onto the *E. coli* chromosome. Whilst this might be a relatively rare event, repeating the above experiments in a $RecA^-$ background would confirm beyond doubt that the pDSA series plasmids express functional C22 and F52a polypeptides.

TABLE 3.10 COMPLEMENTATION OF DSA 103 CHP SENSITIVITY

Strain	Plasmid	Kill Zone (mm)[b]	
		LB	Minimal Medium
A19	-	18.33	20
DSA 103		25	34.5
DSA 103	pDSA 23 $C^+ F^+ Cm^R$	14.5	18.5
DSA 103	pDSA 24 $C^+ Cm^R$	27	33
DSA 103	pDSA 26 $F^+ Ap^R$	+ IPTG[a] 31	38
		- IPTG 29	34
DSA 103	pDSA 24 $C^+ Cm^R$ + pDSA 26 $F^+ Ap^R$	+ IPTG[a] 22.5	36
		- IPTG 19	26
DSA 103	pACYC184	25	34
	pSKS106	24.5	35

a) IPTG included in bottom agar at a final concentration of 3ug ml^{-1}
b) Diameter of the zone of killing in millimeters (figures stated are the mean of three separate measurements).

Discussion

In this chapter, the mapping and characterisation of the gene encoding SCRP-23 has been described. It is 1 of 4 major proteins found to cross-react with antisera raised against a sub-region 2.2 peptide of sigma 70 and sigma 32. From its DNA sequence, the gene was identified as *ahpC*, encoding the C22 sub-unit of alkylhydroperoxide reductase. A downstream reading frame was identified as the *ahpF* gene encoding the F52a sub-unit of the same enzyme, which is thought to rid the cell of lipid hydroperoxides and other toxic oxidative agents.

Storz *et al* (1989) mapped *ahp* by Tn*10* mediated deletions in an *E. coli* K12 strain carrying *zbe-279::Tn10* which co-transduces with *rna* near 14 minutes on the standard *E. coli* linkage map, and found it to be linked to *entA* near 13.85 minutes. In this study the position of the *ahp* locus was determined to be at 653kb on the physical map of the *E. coli* K12 chromosome (see Section 3.2.2), which corresponds to ~14.2 minutes on the genetic map. Moreover an *ahp::kan* marker was found to co-transduce at ~75% frequency with *rna*, which maps near 659kb or ~14.4 minutes (Meador III and Kendal 1990). These results are consistent with the earlier mapping of Storz *et al* (1989) and place *ahp* between *entA* (645kb) and *rna*. The *ahp* locus of *S. typhimurium* similarly maps between the *ent* locus and *lip*, near 13 minutes on the standard linkage map (Storz 1989).

To determine the orientation of the *ahpC* and *F* genes on the chromosome, the order of restriction sites in the determined *ahp* nucleotide sequence was compared with the restriction pattern near the 653kb region of the physical map. From this it is clear that

the *ahp* operon is transcribed clockwise on the *E. coli* K12 chromosome and away from *OriC*. The latter orientation is often associated with highly expressed *E. coli* genes (Brewer 1988), and is consistent with the high frequency of optimal codons within the *ahpC* coding sequence (91%) and the apparent high expression of the C22 and F52a proteins both by lambda recombinants and plasmids, in UV irradiated *E. coli* and minicells respectively.

The *ahp* locus of *S. typhimurium* has been cloned (Storz *et al* 1989) and sequenced (Tartaglia *et al* 1990). The *ahpF* gene is encoded 249 base pairs downstream of the *ahpC* gene and they are probably co-transcribed since two independently isolated Tn*10* insertions **both** abolished expression of the C22 **and** F52a polypeptides. I found the same gene organisation for the *ahpC* and *F* genes in *E. coli*. Moreover the sequence revealed no likely promoters between *ahpC* and *F*. Although such negative evidence is indecisive, it suggests that the genes are also co-transcribed in *E. coli* and raises a question as to the function of the large intergenic region.

The *ahp* operons of *E. coli* and *S. typhimurium* are highly conserved at the nucleotide sequence level, reflecting the close functional similarity which was clearly demonstrated by reciprocal complementation of *E. coli* and *S. typhimurium oxyR* deletion strains by the cloned *ahp* genes (Storz *et al* 1989). Jacobson *et al* have also found the purified proteins to be biochemically similar. The frameshifts apparent upon comparison of the *ahpC* nucleotide sequences of *E. coli* and *S. typhimurium* would, if real, cause the respective C22 polypeptides to differ significantly over one-third of their

C-terminal aa sequences. Therefore it is most likely that the frameshift differences reflect sequencing errors.

It is proposed that the frameshift errors are within the published *S. typhimurium ahpC* sequence for the following reasons:-

The C22 polypeptide is strongly expressed in *E. coli* (see Sections 3.4.1 and 3.4.2) which is reflected in the strong bias towards codons recognised by major iso-accepting tRNA species of *E. coli* apparent within the *ahpC* coding sequence (termed frame II in Section 3.5.4). A similar bias towards codons recognised by major tRNA species is only apparent within the S. typhimurium *ahp* sequence after the reading frame is "corrected" to frame II by insertions of C and G at positions 523 and 688 respectively.

In contrast, the codon usage of the published *ahpC* sequence of *S. typhimurium* (Frame I) as reflected by its Fop of 0.66, is more characteristic of weakly or moderately expressed genes (Ikemura 1985).

Secondly, the polypeptides encoded by Frame II in both organisms are highly homologous (sharing 97.9% identity) and show strong sequence similarity with at least 12 proteins from diverse organisms, throughout their entire lengths. The C22 polypeptides encoded by Frame I would be different in size (due to the occurence of a stop codon (codon 160) within the *E. coli* gene) and would show identity to the related group of proteins only within their N-terminal aa sequences. Storz *et al* have themselves concluded that they made both frameshift errors (personal communication from G. Storz).

The region least conserved between the two organisms is the intergenic sequence between *ahpC* and *ahpF*, which is only 79.6% identical. In *E. coli* it is 4bp shorter, due to 9 insertions and/or deletions, and has several nucleotide substitutions. Whilst this is not surprising, as there is generally less selection pressure upon non-coding DNA sequences, it may be of functional significance. The F52a polypeptide is induced by heatshock in *S. typhimurium* in a manner that is dependent on the OxyR transcription regulator but independent of sigma 32 (Storz *et al* 1989). The C22 polypeptide is not heat inducible in either *E. coli* or *S. typhimurium*. This raises the possibility that heat-induction of the F52a polypeptide may be mediated by an OxyR dependent promoter within the *ahpC-F* intergenic region of *S. typhimurium*. The substantial variation in the DNA sequence of the *ahp* intergenic regions between the two organisms might explain why the OxyR dependent heat induction of the F52a protein observed in *S. typhimurium* but **not in** *E. coli*, if the proposed OxyR dependent promoter upstream of *ahpF* had been lost through mutations in the *E. coli ahp* intergenic region.

An alternative explanation is that heat-shock stimulates OxyR dependent transcription from $ahpP_1$ in both organisms. This would require some form of post transcriptional regulation to effect the observed increase in *S. typhimurium* F52a levels whilst preventing increased expression of the C22 proteins and *E. coli* F52a. Such a mechanism might, in part, be mediated by the different *ahpC-F* intergenic sequences in *E. coli* and *S. typhimurium*.
The mechanism by which OxyR is activated by heat is currently unknown. However it is thought that increased temperatures may

produce higher levels of intracellular active oxygen species (reviewed by Demple 1991) thereby causing oxidative stress.

Obviously it is desirable to test the above speculations by transcriptional analysis of the *ahp* intergenic region. Primer extension analysis and/or S1 nuclease mapping of *ahp* mRNA from wild type and OxyR deletion strains of *S. typhimurium*, grown at 30°C or 42°C, should determine the start site of any OxyR dependent transcripts initiated upstream of *ahpF* upon heat-shock. A similar analysis of *ahpF transcription in E. coli* should provide comparative data and confirm whether the *ahpC* and *ahpF* genes are co-transcribed.

Alternatively mRNA hybridisation analyses might reveal that both *ahpC* and *F* are transcribed at elevated levels after heat-shock implying differential post-transcriptional effects on *ahpC* and *F* expression.

OxyR activates transcription of several peroxide stress genes by binding to a site upstream of and overlapping the -35 promoter element (Tartaglia *et al* 1989, 1992; Storz *et al* 1990; Tao *et al* 1991) and it autorepresses *oxyR* transcription by sterically blocking RNA polymerase access to the promoter (Tao *et al* 1991).

Storz *et al* (1990) and Tartaglia *et al* (1992) suggested that OxyR is permanently bound at dependent promoters, where it is poised to activate transcription upon its oxidation by H_2O_2 or other oxidising agents. This OxyR activation model was based on DNaseI footprinting data which showed that *E. coli* OxyR binds tightly to its dependent promoters *in vitro* both in air saturated buffers (where it is active) and under highly reducing conditions (when OxyR is inactive).

The results presented in Section 3.6.2 show that in *E. coli ahpC* is transcribed from at least 2 promoters. Of these, $ahpP_1$ is clearly induced by H_2O_2 *in vivo* and activated by OxyR *in vitro*, whilst $ahpP_2$ is OxyR independent and is repressed *in vivo* by H_2O_2 treatment and *in vitro* by oxidised (active) OxyR. $ahpP_2$ is located within the putative OxyR target upstream of $ahpP_1$. Therefore it was initially surprising to find that $ahpP_2$ is significantly active *in vivo* under normal aerobic conditions, as OxyR should bind to a site extending between bp-36 and bp+7 relative to the P_2 transcription start site. Indeed OxyR represses *oxyR* transcription by binding in a fairly similar position (beween bp-26 and bp+18). If OxyR were permanently bound at the *ahp* promoter then transcription from $ahpP_2$ should be repressed at all times. The pattern of *ahp* transcription actually observed in *E. coli* is more consistent with a model whereby only oxidised active OxyR can recognise and bind to the *ahp* promoter region.

An analagous pattern of transcription has been reported for the *E. coli katG* gene, which is also transcribed from dual promoters *in vivo* (Tao *et al* 1991). A second promoter has been mapped within the experimentally determined OxyR binding site upstream of the *katG* gene, which is active *in vivo* under normal aerobic growth conditions but repressed within $OxyR^{(cons)}$ strains. Transcription from this promoter is also repressed by active OxyR *in vitro* (Tao *et al* 1993). These results appear to conflict with those of Storz *et al* 1990 and Tartaglia *et al* 1992, who found that OxyR bound tightly to its dependent promoters *in vitro* even under highly reducing conditions, namely in buffers containing 100mM DTT. However, the interpretation

of their data assumed that the *in vitro* conditions employed accurately reflected the intracellular reducing environment, and that there were no significant amounts of active (oxidised) OxyR in the presence of 100mM DTT. However recently it was found that the Cys 199-Ser- OxyR mutant, which is permanently "locked" in the reduced state, does not bind to the *E. coli katG* promoter nor the *S. typhimurium ahpC* promoter (personal communication from G. Storz). These results are consistent with the data presented here, providing strong evidence that reduced OxyR does not bind to the *katG* and *ahp* promoters. Furthermore, when considered with previous data which found that OxyR Cys-199-Ser will bind to and repress transcription of the *oxyR* gene (Storz *et al* 1990), it appears that the recognition properties of OxyR are different, depending on whether it is acting as auto-repressor or as an activator of *ahp* and *katG*. Indeed, the contacts made by OxyR upstream of the *ahp* and *katG* promoters are different from those made at the *oxyR* promoter (personal communication from G. Storz).

The function of $ahpP_2$ is unclear. It is rather weak and may be an evolutionary relic, although it probably contributes to basal level expression of *ahp*. The analagous promoter is apparent within the DNA sequence upstream of the *ahp* gene in *S. typhimurium* (see figure 3.11), differing only at position 5 of the -35 sequence, but is inactive even during normal aerobic growth (Tartaglia *et al* 1990). $ahpP_2$ may be required for *ahp* transcription under unknown stress conditions, possibly requiring accessory factors for transcriptional activation. The $ahpP_2$ promoter might then be active in *S. typhimurium* as well as *E. coli*.

Recently, in unpublished work, Francis *et al* observed that light emission from an *S. typhimurium* strain carrying an *ahp::lux* transcriptional fusion is elevated upon entry into stationary phase, suggesting a possible role for $ahpP_2$ in the stationary phase response. Another gene which has recently been found to be OxyR dependent - *dps* - (personal communication from G. Storz) is also induced during stationary phase by sigma S, as are other peroxide resistance functions; namely HPII catalase (encoded by *katE*) and Exonuclease III (encoded by *xtha*) which are <u>not</u> OxyR dependent (reviewed by Hengge-Aronis 1993). Sigma S is responsible for the transcription of a large subset of genes during stationary phase and it is possible that *ahp* transcription during stationary phase is sigma S dependent. Obviously more detailed transcriptional data would be required to verify this and indeed define the *ahp* promoter(s) responsible for stationary phase transcription.

The *E. coli* and *S. typhimurium* C22 polypeptides are members of a large group of proteins from diverse organisms which share pairwise sequence identities ranging from 97.9% to 13.5%. Whilst some of the overall pairwise identities listed in Table 3.6 are rather low, the clustering of identical amino acids within 4 common regions strongly suggests that the C22 group proteins may share higher order structural and perhaps functional features. As yet there are insufficient functional data to verify this and it would be premature to describe the C22 group as a new protein family. For convenience therefore the C22 group has been defined as a protein clan, that is, a group of proteins which share common primary structural features suggesting that they may be related, but for which a common

functional relationship has yet to be demonstrated.

The most striking region of similarity between the C22 clan proteins is Region 1. It includes an invariant Cys residue, which is suspected to play a role in redox reactions in the *E. coli* and *S. typhimurium* C22 proteins. Flores *et al* (1993) noted the similarity in Region 1 between a subgroup of C22 clan proteins (ahpC_ salty, YNDH_BACSP, MER5_Mouse, 26kD_HELPY, R20K_CLOPA and CR29_ENTHI in Table 3.4), during their investigation of the major free-thiol containing surface protein of *Entamoeba histolytica* (CR29_ENTHI). The 28kDa protein is required for survival of the parasite in extra-intestinal aerobic environments (Gillin *et al* 1984), possibly through the action of redox active cysteines (Gillin and Diamond 1981 a & b).

Flores *et al* (1993) proposed that Region 1 may represent a common functional domain of the C22 clan proteins. To ascertain whether such sequences also occur outside the clan, the SWISS-PROT and NBRF-PIR databases were searched for sequences matching Region 1. Only known C22 clan proteins showed significant similarity with this Region, indicating that it may represent a characteristic "signature" of C22 clan proteins which could be used in future identification of new clan members.

Within the C22 clan there are sub-groups of proteins which might represent true families, based upon strong sequence similarities and comparisons with systems of known function. The strongest candidate family encompasses the *E. coli* and *S. typhimurium* C22 proteins and putative polypeptides encoded by open reading frames located shortly upstream of the *ndh* and *nox* genes of *B. alcalophilus* and *A. xylanus*

respectively (the sequences designated YNDH_ BACSP and AXYNOX respectively, in table 3.4). The *ndh* and *nox* genes encode an NADH dehydrogenase (NDH) and an NADH Oxidase (NOX) respectively, which are closely related in sequence (including known functional domains) to the F52a polypeptide of Ahp and to thioredoxin reductase (discussed in full below). The C22 polypeptides and the products of the upstream *orfs* within the YNDH_BACSP and AXYNOX sequences share pairwise identities between 97.8% and 64.3% (see Table 3.6), which when considered along with their chromosomal locations strongly suggests that they may be functionally related. Claiborne *et al* (1992) have suggested (on comparison with the thioredoxin - thioredoxin reductase system) that the F52a polypeptide of Ahp transfers electrons from NADPH, via a redox active disulphide and an FAD cofactor, to a second redox active disulphide on the C22 protein which then reduces the substrate alkylhydroperoxide. NDH is widely believed to be the respiratory dehydrogenase of *B. alcalophilus* (Xu *et al* 1991). By analogy with Ahp, NDH might transfer electrons to the C22 clan protein within the *B. alcalophilus* respiratory chain. It should be noted that the respiratory NADH dehydrogenase of *E. coli* is quite different, and probably evolved from a different flavin-containing ancestor than did the F52a polypeptide and NDH (Xu *et al* 1991). An alternative but less likely possibility is that the NDH and C22 clan proteins of *B. alcalophilus* represent the Ahp of that organism. The C22 clan protein and NOX of *A. xylanus* may similarly function as redox partners, but in this case there is no respiratory chain to which they could be assigned.

A possible fifth member of the C22 family is the product of orf 3

within the rubredoxin operon of *C. pasteurianum* (Mathieu *et al* 1992). It is less similar to the above four proteins than they are to each other, sharing a maximal identity of 34.3% with the C22 proteins of *E. coli* and *S. typhimurium*. However it is transcribed from the same operon as *orf 1*, which encodes a polypeptide strongly related in sequence to thioredoxin reductase and therefore related to the *E. coli* F52a protein. Rubredoxin itself is a small non-haem iron-containing protein which participates in various electron transport systems in a number of different bacteria. One plausible hypothesis is that the products of *orf* 1 and *orf* 3 (encoded in the reverse order to the *ahpC* and *F* genes) represent a rubredoxin oxidoreductase. Rubredoxin NADH oxidoreductases have been partially purified from *C. acetobutylicum* and *C. sticklandii* although their subunit structures were not reported. (Stadtman 1965, Ballangue *et al* 1986). Obviously functional characterisation of the *orf* 1 and *orf 3* gene products would be required to verify the above speculation.

Of the remaining C22 clan proteins, only the 29kDa surface antigen of *E. histolytica* has been implicated as a redox active protein. However the strong aa sequence similarities surrounding the invariant Cys residues within Regions 1 and 4 of other core group proteins suggest that these may also be redox active proteins. Some of the source organisms - *E. histolytica, H. pylori, M. avium* and *M. leprae* are pathogens which may encounter the oxidative threat of extra-intestinal environments, or the oxidative attack mechanisms of macrophages. It is possible that the C22 clan proteins of these organisms play a similar protective role to that of C22 in *E. coli* and *S. typhimurium*. The relationship between the Mer5 protein of

Mouse erythroleukaemic cells and the C22 protein is more puzzling. Mer5 is elevated within erythroleukaemic cells induced to differentiate, *in vitro*, with DMSO (Nemoto *et al* 1990). Its role in differentiation is essential, as Mer5 antisense RNA blocks differentiation when expressed within erythroleukaemic cells. There is no currently known connection between Mer5 function and oxidative defence proteins. However as Flores *et al* (1993) have noted, erythroleukaemic cells must evade immune defences in order to cause disease. Obviously these defences might include the oxidative agents released by macrophages. Therefore further functional studies of the Mer5 protein might usefully ask whether it is involved in defence against oxidative stress.

The three non-core group proteins within the C22 clan (figure 3.13) show only poor similarity within region 4.2 and lack its Cys residue, which is invariant in the core group. Additionally BCP of *E. coli*, which is the least similar protein in the alignment, lacks sub-region 4.1 completely. MBF_SOD is encoded upstream of the superoxide dismutase of *M. thermoautotrophicum*. Whether these two juxtaposed genes are parts of an oxidative stress operon cannot be determined without more direct functional information.

There is no evidence that either BCP of *E. coli* or BMU_EMSP (the polypeptide product of an embryo-specific transcript of the grass *Bromus seccalinus*) are involved in electron transfer reactions or defences against oxidative stress. BMU_EMSP is expressed at elevated levels within dormant seeds as opposed to non-dormant seeds. Defence against oxidation may be particularly important in dormant

cells, suggesting a very tentative speculation as to the role of BMU_EMSP.

As described in Section 3.7.2 the F52a protein of Ahp is a member of the thioredoxin reductase family of flavoprotein disulphide oxido-reductases. The strong similarity between the F52a protein and thioredoxin reductase has been discussed previously (Tartaglia *et al* 1990). Similarly Xu *et al* (1991) and Mathieu *et al* (1992) have noted the similarity between NDH of *B. alcalophilus* and thioredoxin reductase, and between the *C. pasteurianum* orf 3 gene product and F52a and thioredoxin reductase, respectively. All members of this group show perfect conservation of a pair of Cys residues, which are redox active in thioredoxin reductase, and their positions relative to a proposed NAD(P) binding sequence, whilst two FAD cofactor binding sequences are also conserved both in sequence and position.

A significant new piece of information apparent in Figure 3.14 is the finding that the N-termini of the F52a proteins of *E. coli* and *S. typhimurium* show strong sequence similarity to the membrane binding domain of NDH. In *B. alcalophilus* this domain serves to attach NDH loosely to the cell membrane and indeed can be proteolytically cleaved from the C-terminal catalytic domain without affecting the function of either component (Xu *et al* 1989).

The occurrence of similar sequences within the F52a protein and NOX of *A. xylanus* (also recently noted by Niimura *et al* 1993) suggests that these proteins may also associate loosely with the cell membrane via a similar N-terminal domain. Whilst this needs confirmation, the localisation of the F52a protein to the cell membrane would be

consistent with its proposed function in the elimination of lipid hydroperoxides, which pose a substantial oxidative threat to membrane function and integrity.

Jacobson *et al* (1989) purified the Ahp sub-units from *S. typhimurium* and *E. coli* and found that the C22 protein was present at substantially higher levels than the F52a protein in crude cell lysates. Expression of the Ahp proteins was monitored in UV irradiated *E. coli* infected with lambda clones and in purified minicells bearing pDSA23. In both systems the C22 protein appears to have been synthesised in a large molar excess over the F52a protein. The functional significance of this is unclear. However, the C22 polypeptide behaves similarly to the 29kDa surface antigen of *E. histolytica* in that both polypeptides form high molecular weight aggregates *in vitro*. Moreover ~200kDa aggregates of the 29kDa protein have been detected within the *E. histolytica* membrane and are thought to be the major functional form of the protein (Flores *et al* 1993). At this time there is no direct evidence regarding the functional sub-unit structure of the C22 protein, so that the explanation of the relative amounts of C22 and F52a proteins within the cell is open to speculation. Tartaglia *et al* (1990) suggested that the C22 protein might target and bind the alkylhydroperoxide substrate. Therefore it is conceivable that the C22 protein may play a sequestering role to "mop up" toxic alkylhydroperoxides. If this is true, then individual C22 sub-units may only transiently associate with F52a dimers, for as long as is necessary to receive electrons for subsequent substrate reduction. However, the apparent co-precipitation of the F52a protein along with the C22 protein by

anti-2.2 peptide-antiserum (see Section 3.4.3) suggests that at any given time a significant number of F52a dimers are associated with C22 proteins. (The F52a polypeptide exists in solution as a homo-dimer, which is thought to be its functionally relevant structure: Jacobson *et al* 1989).

As discussed above, the data presented in this chapter have raised several questions concerning *ahp* regulation and function. The *ahp*::*Kan* strain DSA104 and the plasmids pDSA23, pDSA24 and pDSA26 are starting materials which should allow some of these questions to be addressed.

The data presented in Table 3.8 show that pDSA23 will complement the CHP sensitivity of DSA103, as will pDSA24 and pDSA26 in combination, indicating that all three plasmids express functional C22 and/or F52a proteins. In addition pDSA23 will complement the CHP sensitivity of an *S. typhimurium* strain which carries an *ahpC*::*lux* transcriptional fusion, (Francis and Gallacher unpublished) which is entirely consistent with the previous observation of Storz *et al* (1989) that the cloned *E. coli ahp* operon will complement the CHP sensitivity of a *S. typhimurium ahp* strain.

An obvious question which needs to be addressed is the role of the conserved regions, 1-4, in Ahp structure and function, in particular the invariant Cys residues - Cys 49 and Cys 166 - within regions 1 and 4, which are likely to form the redox active disulphide predicted by the chemical evidence of Jacobson *et al* (1989). Site directed mutagenesis, targetting either or both of Cys 49 and Cys 166 should

prove useful in evaluating the redox properties of the C22 protein. Several other conserved residues, including the clustered aromatic residues in Region 1, are obvious targets for directed mutagenesis, whilst more random mutagenesis of the *ahpC* gene might pinpoint other important residues which participate in catalysis of alkylhydroperoxide reduction. To this end, DSA103 should prove a useful Ahp^- host for *in vivo* analysis of mutant C22 proteins, whilst the Ahp^- genetic background may also aid their purification free from "contamination" by wild type protein.

Similar analysis of the redox properties of the F52a protein may also prove useful, and indeed would provide comparative data for those generated by Prongay *et al* (1992) who have made Cys-135-Ser and Cys-138-Ser substitutions in thioredoxin reductase to assess the roles of these Cys residues in the redox active disulphide. Cys 138 has been identified as the residue which receives electrons from $FADH_2$ in catalysis. By analogy it might be expected that Cys 349, the corresponding residue of the F52a protein might function in a similar way.

Additionally, the possible membrane binding function of the N-terminal domain of the F52a protein is worth investigating. This was achieved for NDH by assaying its ability to attach to liposomes of similar lipid composition to the *B. alcalophilus* cell membrane. pDSA26 should allow expression of the F52a protein or mutant derivatives, for membrane binding studies similar to those described for NDH by Xu *et al* (1989). These might initially involve deletion analysis of the F52a protein N-terminus, followed by more targetted

point mutagenesis in order to identify sequences responsible for membrane attachment. The importance of membrane localisation to Ahp function might also be assayed *in vivo* using the same materials, by measuring the ability of the cloned *ahp* genes to complement an *ahp* deletion strain.

CHAPTER 4

4.1 INTRODUCTION

An SCRP of 27 kDa, initially identified by Fujita *et al* (1987), was subsequently found to be a mixture of two proteins both of which cross-react specifically with antibodies in the polyclonal mixture raised against the sigma 70-2.2 peptide (Ueshima *et al* 1992). One of these was SCRPs was found to be ribosomal protein S2. The other, SCRP-27A, which migrates marginally slowly than S2 in high resolution SDS-PAGE, was a previously unkown protein in view of its N-terminal amino acid sequence, shown below:

MKKIGVILSGXGVYDGSEIHEAVL (X = unknown)

SCRPS of 27 kDa were found to be distributed throughout all fractions, including RNAP holoenzyme fractions, after separation of crude *E. coli* cell extracts in glycerol gradients. However the individual behaviour of the two SCRP 27 proteins in such gradients was not examined, although it was shown that neither associates with RNAP as judged by fractionation of mixtures of purified 27 kDa SCRPS and RNAP core enzyme on glycerol gradients. It is therefore possible that either SCRP-27A or S2 or both associate with other cellular components.

In order to derive more information concerning the nature of SCRP-27A, the corresponding gene, provisionally named *scrP*, was cloned and sequenced. Part of this work was carried out in collaboration with an undergraduate student,

Fiona M. Townsley, who worked for a short time under my supervision in this laboratory. Work done by her is acknowledged in the appropriate figure legends.

4.2.1 Physical Mapping of the *scrp27A* locus on the *E. coli* K-12 Chromosome

Rough mapping of the *scrP* gene had been carried out independently by Taeko Suzuki and Dr. Richard Hayward at the National Institute of Genetics, Mishima, Japan. A pair of degenerate oligo deoxyribonucleotides (*scrp*27A-1 [5'-ATGAA(G/A)AA(G/A)AT(T/C/A)GGNGT] {N=A,C,G or T} with a degeneracy of 48, corresponding to aa 1 through 6 of SCRP-27A and *scrp*27A-2 [5'-GA(G/A AT(T/C/A)CA(T/C)GA(G/A)GCN GT] with a degeneracy of 96, corresponding to aa 18 through 23 of SCRP-27A) were labelled at their 5' ends with gamma [^{32}P]ATP, as described in Section 2.8.2 and hybridised to nylon filters carrying DNA from the lambda miniset of *E. coli* K12 genomic clones, as described in Section 2.8.5. Both probes hybridised with a pair of contiguous clones, lambda 522 and lambda 523, which overlap by ~9kb near 3420kb on the physical map (Kohara 1987), corresponding to ~69.5 minutes on the genetic map (Bachman 1990) (Figure 4.1).

4.2.2 Fine Mapping of *scrp*27A

The physical map location of *scrP* was confirmed and more

precisely defined by Restriction and Southern hybridisation analyses of lambda 522 and lambda 523 DNAs (using 5'-[^{32}P] labelled *scrp*27A-2) as described in Sections 2.6.1 and 2.8.5 (lambda 522 and lambda 523 were grown in NM621 and their DNAs purified as described in Sections 2.4.7 and 2.5.5).

Restriction analysis confirmed the identity of lambda 522 and lambda 523 (Figure 4.2), whilst the scrp27-A2-responsive site was found to lie between *Eco*RV and *Pvu*II sites located near 3416 kb and 3417 respectively on the physical map (Figures 4.3 and 4.4). The closest mapped genes on either side are *arcB* near 3420 kb (Iuchi and Lin 1990) and *rpoN* together with 3 associated open reading frames near 3410-3415 kb (Merrick *et al* 1993). The restriction map presented in figure 4.4 is based on restriction analyses and the consensus physical map of Kohara *et al* (1987) and does not differ significantly from the latter map.

Figure 4.1

The locations of the *E. coli* DNA inserts in lambda 522 and lambda 523 on the physical map. The top scale shows co-ordinates on the E. coli K12 physical map (Kohara *et al* 1987) in kilobases. The extent of E. coli chromosomal DNA fragments carried by the miniset clones (*ibid*) is indicated by solid lines under the scale; the identity of each individual clone is shown. The positions of mapped markers are shown above the scale.

FIGURE 4.1

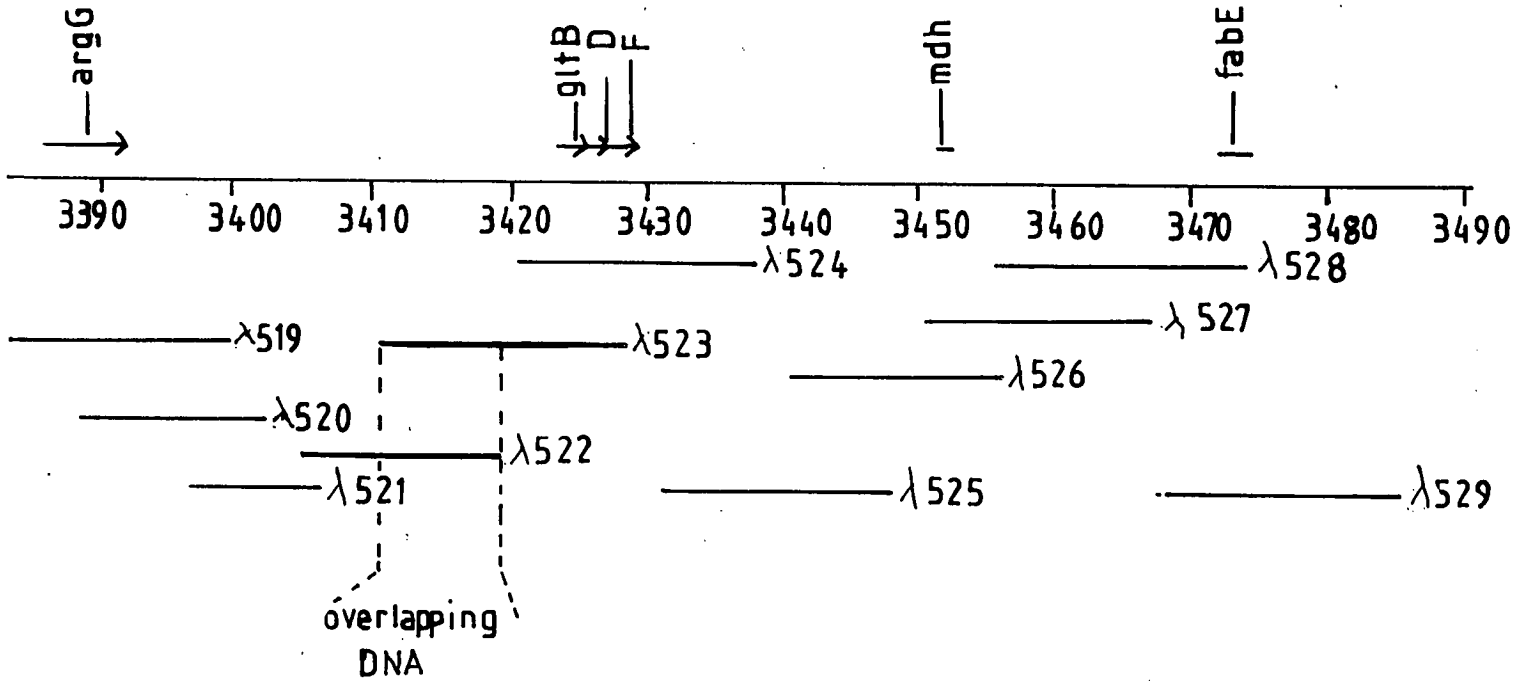

argG
gltB
D
F
mdh
fabE
3390
3400
3410
3420
3430
3440
3450
3460
3470
3480
3490
λ524
λ528
λ527
λ519
λ523
λ526
λ520
λ522
λ521
λ525
λ529
overlapping
DNA

Figure 4.2

Restriction analysis of lambda 522 and lambda 523. 0.6% agarose gel. Tracks 1, 3, 5, 8, 10 and 12 show lambda 522 DNA; Tracks 2, 4, 6, 9, 11 and 13 show lambda 523 DNA and Tracks 7 and 14 show lambda *c*I857 *S*am7 marker DNA digested with *Hind*III + *Eco*RI and *Hind*III respectively. The relevant marker bands are indicated and their sizes given (in base pairs) to the right of the gel photograph. The following abbreviations are used for restriction enzymes: B, *Bam*HI; E, *Eco*RV; H, *Hind*III; K, *Kpn*I; R, *Eco*RI and V, *Pvu*II.

FIGURE 4.2

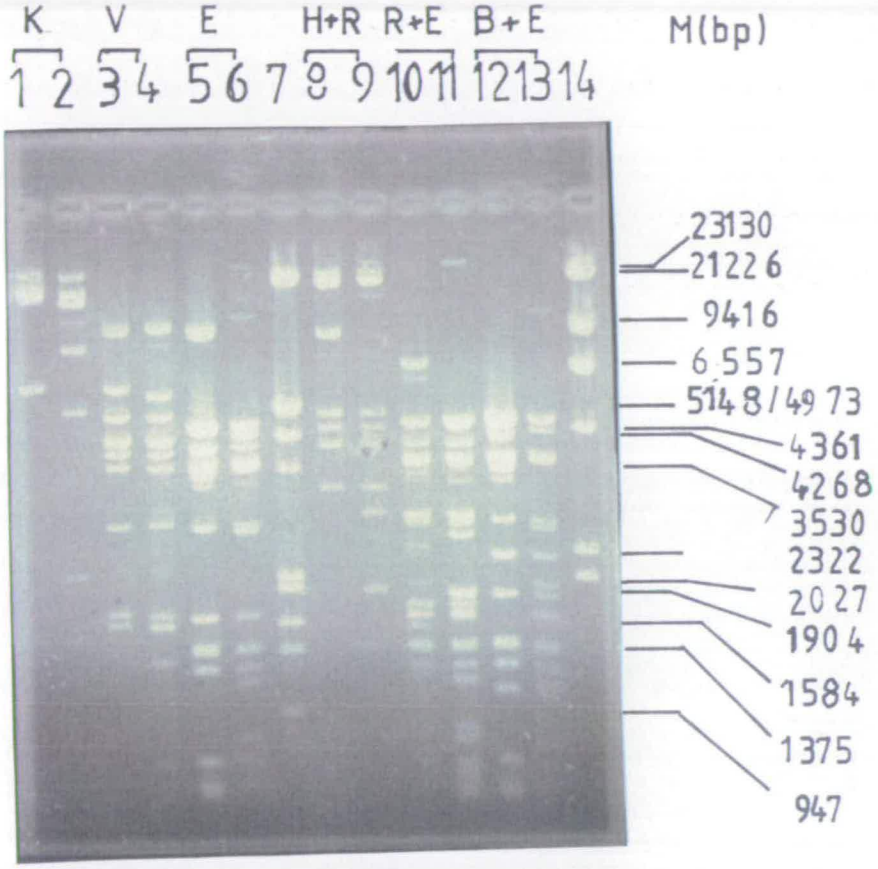

K V E H+R R+E B+E M(bp)
1 2 3 4 5 6 7 8 9 10 11 12 13 14
23130
21226
9416
6557
5148/4973
4361
4268
3530
2322
2027
1904
1584
1375
947

Figure 4.3

Fine mapping of the *scrP* gene by Southern blotting using scrp27A-2 as probe. DNA was transferred from the gel shown in Figure 4.2 to a nylon membrane as described in materials and methods. Hybridisation was performed at 44°C for 16 hours. Final washing conditions were 2 x SSC, 0.2% SDS for two periods of 5 minutes at 40°C. Tracks 1, 3, 5, 8, 10 and 12 show lambda 522 DNA; Tracks 2, 4, 6, 9, 11 and 13 show lambda 523 DNA. Tracks 7 and 14 contain lambda *c*I857 *S*am7 DNA digested with *Hin*dIII + *Eco*RI and *Hin*dIII respectively. Relevant marker fragments are indicated and their sizes given in base pairs. Filled arrow heads point to probe positive *Pvu*II (Track 4) and *Eco*RV (Track 5) fragments which define the limits of the smallest positively hybridising site. The following abbreviations are used for restriction enzymes: B, *Bam*HI; E, *Eco*RV; H, *Hin*dIII; K, *Kpn*I; R, *Eco*RI and V, *Pvu*II.

FIGURE 4.3

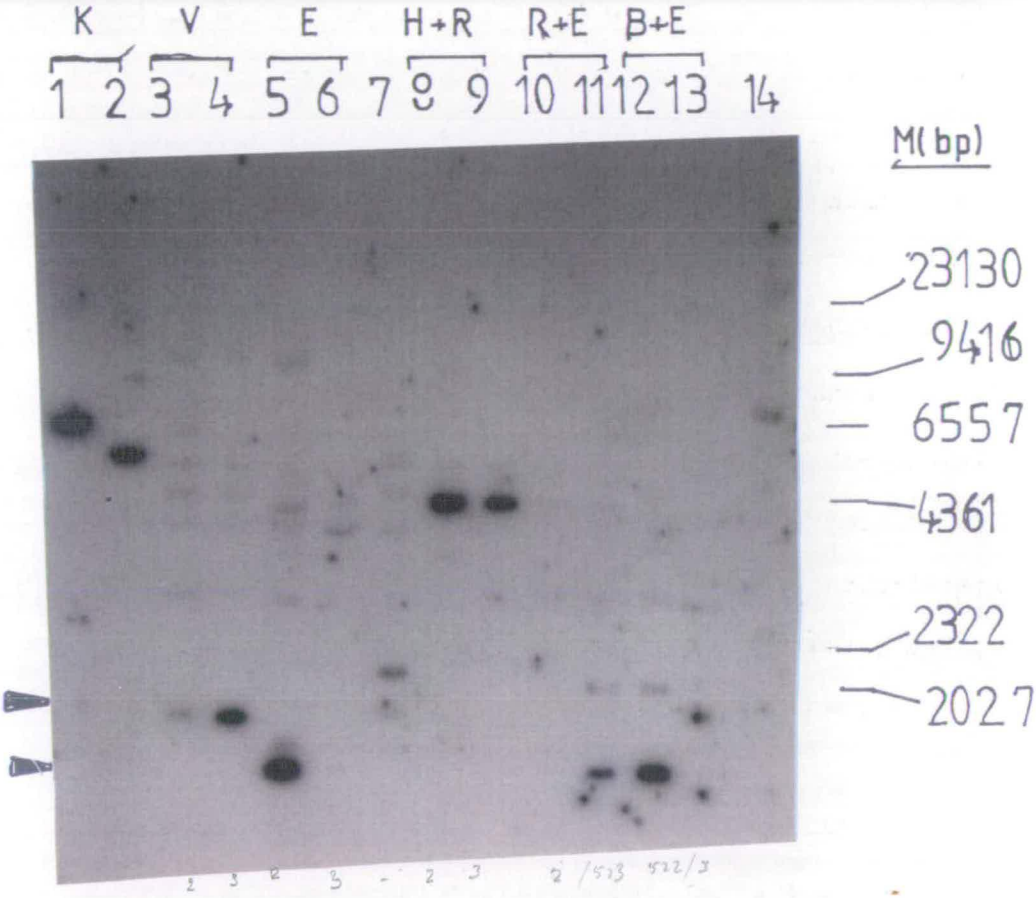

K
V
E
H+R
R+E
B+E
1 2 3 4 5 6 7 8 9 10 11 12 13 14
M(bp)
23130
9416
6557
4361
2322
2027

Figure 4.4

Restriction map of *E. coli* chromosomal inserts cloned in lambdas 522 and 523. This map is based on restriction analyses performed by Fiona M. Townsley and the author, and also the restriction map of Kohara *et al* (1987). The top scale shows co-ordinates on the physical map of the E. coli K12 chromosome (*ibid*). The locations of the *rpoN*, *arcB* and *scrP* genes are indicated; arrows show the 5′ to 3′ orientation where it is known. Probe positive *Eco*RV and *Pvu*II fragments are highlighted in red, and the length and orientation of the chromosomal DNA inserts carried by lambdas 522 and 523 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*I; K, *Kpn*I; P, *Pst*I; R, *Eco*RI and V, *Pvu*II.

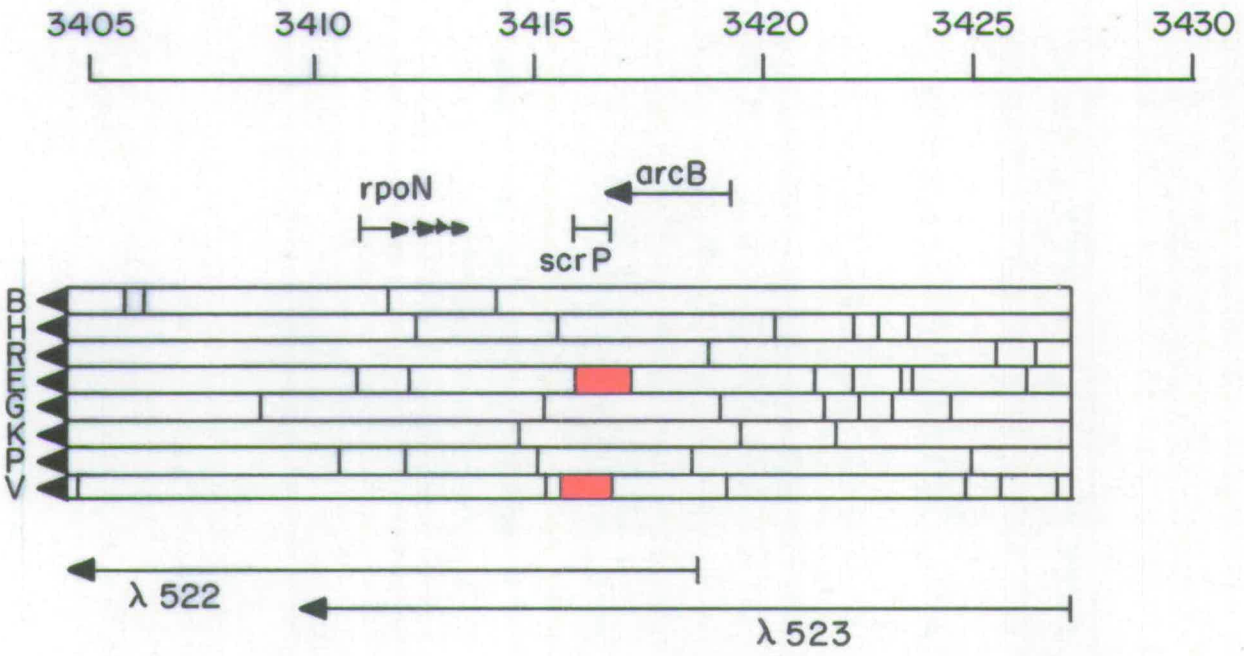

3405
3410
3415
3420
3425
3430
rpoN
arcB
scrP
B
I
R
E
G
K
P
V
λ 522
λ 523

FIGURE 44

4.3 Sub-cloning of the *scrp*27A Locus

In order to characterise the *scrp*27A locus further, M13 and plasmid sub-clones were constructed to facilitate determination of its sequence and identification of its encoded protein(s).

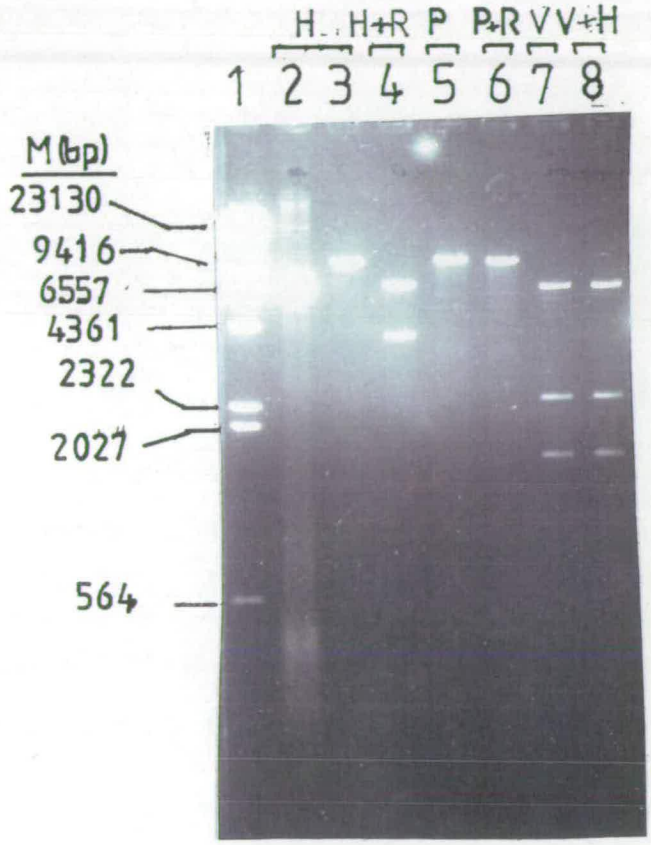
The M13 clone mp18-15 was constructed by inserting the ~4.0 kb *EcoRI-HindIII* fragment of lambda 522 DNA, which hybridised with *scrP*27A-2, between the same sites in mp18. The size and orientation of the mp18-15 insert was checked by restriction analysis (Figure 4.5). M13 mp19-23 (Figure 4.5) was constructed by moving the *Eco*RI-*Hind*III insert of mp18-15 into mp19. The structure of mp19-23 was also checked by restriction analysis (data not shown).

The plasmid sub-clone, pFMT1 was constructed by replacing the 1.09 kb *Eco*RI-*Hind*III fragment of pJW11 with the 4.0 kb *Eco*RI-*Hind*III fragment of lambda 522. The structure of pFMTI was verified by restriction analysis (Figure 4.6).

Figure 4.5

Restriction analysis of mp18-15. 0.8% agarose gel electrophoresis. The top part shows the restriction gel. Track 1 contains lambda *c*I857 *S*am7 marker DNA digested with *Hin*dIII; the relevant bands are marked and their sizes given to the left of the gel photograph. Track 2 contains mp18 DNA; Tracks 3-8 inclusive contain mp18-15 DNA digested with various restriction enzymes as indicated above the gel photograph. The lower part shows partial restriction maps of mp18-15 and mp19-23. Open rectangles represent *arc*B; arrows indicate the 5' to 3' orientation. Unfilled thick lines represent *E. coli* genomic DNA carrying the putative *scr*P gene. M13 DNA is shown as far as the *Pvu*II site (position 5960 in mp18) to the left of the polylinker region, and up to the *Bgl*II site (position 6935 in mp18) to the right. Only relevant restriction sites are shown. The following abbreviations are used for restriction enzymes: H, *Hin*dIII; L, *Bgl*II; P, *Pst*I; R, *Eco*RI and V, *Pvu*II. mp18-15 was constructed by Fiona M. Townsley. Restriction analysis of mp18-15 and the construction of mp19-23 were carried out by the author.

FIGURE 4.5


H
H+R
P
P+R
V
V+H
1 2 3 4 5 6 7 8
M(bp)
23130
9416
6557
4361
2322
2027
564

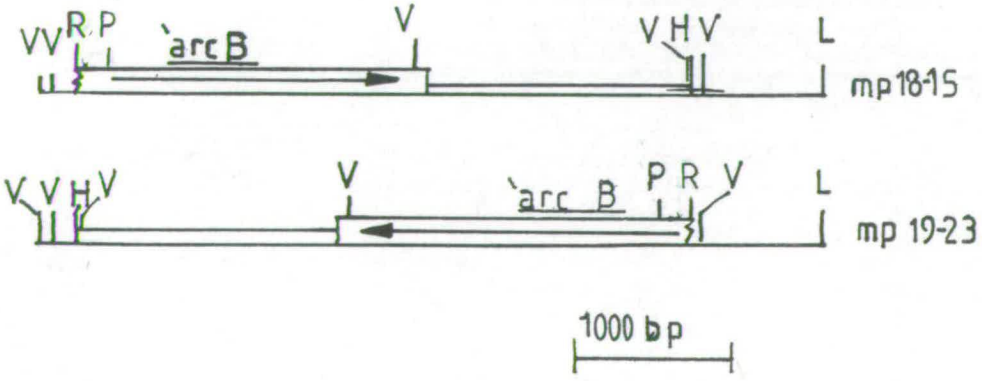
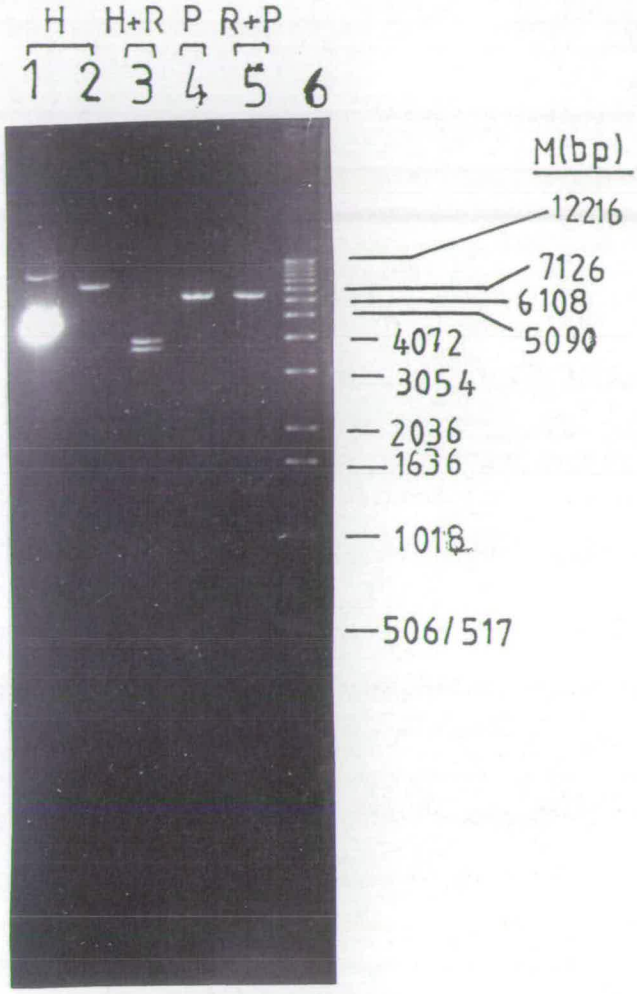
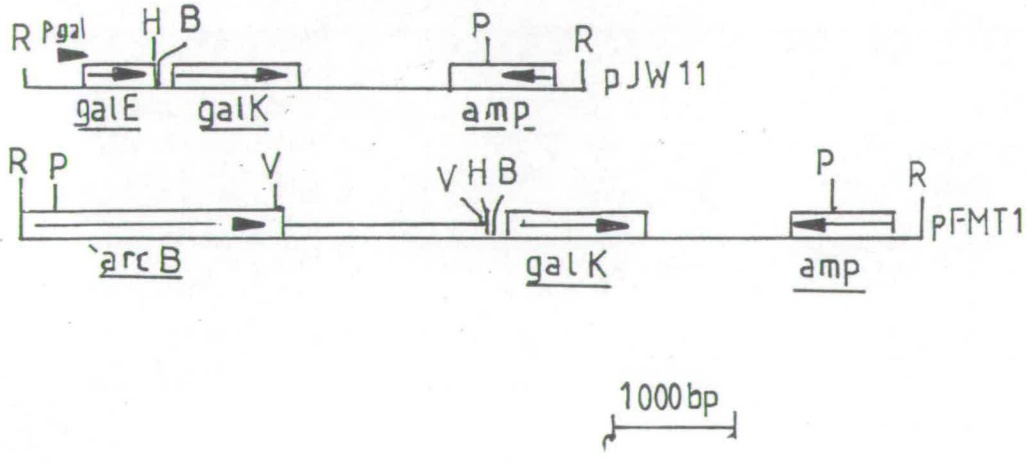

VV
R
P
'arc B
V
V
H
V
L
mp 18-15
V
V
H
V
V
'arc B
P
R
V
L
mp 19-23
1000 bp

Figure 4.6

Restriction analysis of pFMT1. 0.9% agarose gel electrophoresis. Track 6 contains 1 kilobase ladder marker DNA (BRL); the relevant bands are marked and their sizes given to the right of the gel photograph. Track 1 contains pJW11 DNA; Tracks 2-5 inclusive contain pFMT 1 DNA digested with restriction enzymes as indicated above the gel photograph. The lower part shows linear restriction maps of pJW11 and pFMT1; both plasmids are shown linearised at the unique *Eco*RI site position 0 in pJW11. Open rectangles within the plasmid outlines represent relevant genes: *amp* encodes beta-lactamase; *galK* encodes galactokinase and *galE* encodes C-terminally truncated UDP-galactose 4-epimerase. The 5′ to 3′ orientations of the respective genes are indicated by arrows. Only the relevant restriction sites are shown. The following abbreviations are used for restriction enzymes: B, *Bam*HI; H, *Hind*III; P, *Pst*I; R, *Eco*RI and V, *Pvu*II. pFMT1 was constructed by Fiona M. Townsley; restriction analysis was carried out by the author.

FIGURE 4.6


H
H+R
P
R+P
1
2
3
4
5
6
M(bp)
12216
7126
6108
5090
4072
3054
2036
1636
1018
506/517


R
Pgal
H
B
P
R
pJW11
galE
galK
amp
R
P
V
V
H
B
P
R
pFMT1
arcB
galK
amp
1000bp

4.4 Direction of Transcription of *scrP* on the *E. coli* K12 Chromosome

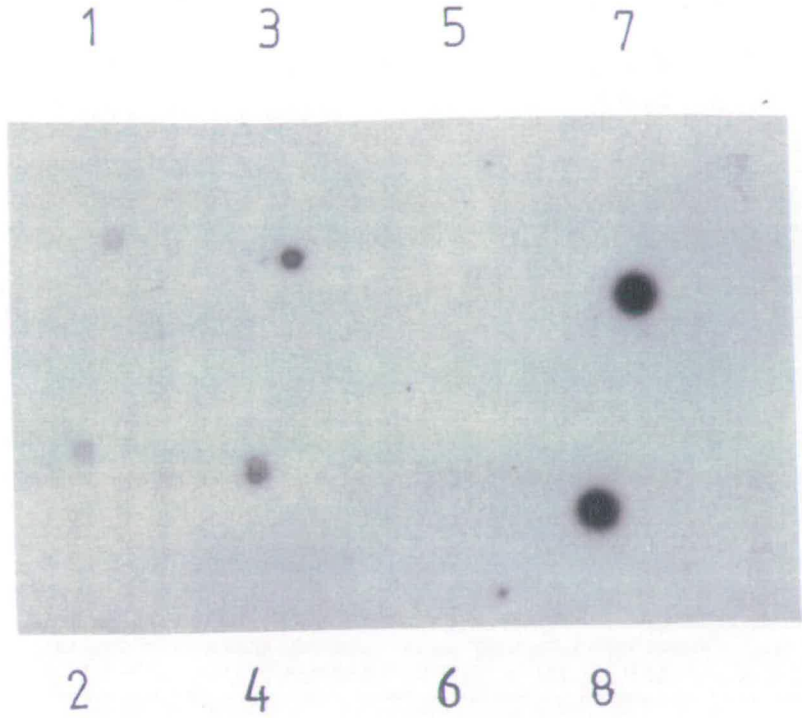
The orientation of *scrP* on the *E. coli* K12 chromosome was determined by dot blot hybridisations using the single stranded forms of mp18-15 and mp19-23 DNAs, with 5'-[^{32}P] labelled *scrp*27A-1 as probe, as described in Section 2.8.2.

As *scrp*27A-1 has the same "sense" as the mRNA-like strand of *scrP*, its positive hybridisation with mp19-23 but **not** mp18-15 (Figure 4.7) predicts that *scrP* is transcribed in the direction from the *Eco*RI site towards the *Hind*III site in mp18-15 and hence counterclockwise on the *E. coli* K12 chromosome. Therefore *scrP* is transcribed in the same direction as *arcB* which lies near to *scrP* on the genetic map (Iuchi and Lin 1990).

Figure 4.7

Dot blot hybridisation analysis of mp18-15 and mp19-23 using scrp27A-1 as probe. ~100ng of ssDNA or ~200ng of RF DNA was incubated at 95°C for 5 minutes then spotted onto a nylon membrane which was then prepared for hybridisation as described in sections 2.8.3 and 2.8.4. Hybridisation was performed for 16 hours at 44°C, and final washing at 40°C in 2 x SSC, 0.2% SDS for two periods of 5 minutes. 1 and 2 contain mp18 RF DNA; 3 and 4 contain mp18-15 RF DNA; 5 and 6 contain mp18-15 ssDNA and 7 and 8 contain mp19-23 ssDNA.

FIGURE 4.7



4.5 Synthesis of pFMT1 Encoded Proteins *In vivo*

SCRP-27A is evidently produced by exponentially growing *E. coli* cells, and cross reacts strongly with anti-2.2-peptide-antibodies on Western blots (Fujita *et al* 1987, Ueshima *et al* 1992). The polypeptides encoded by the *Hin*dIII-*Eco*RI insert of pFMT-1 were examined in an attempt to determine whether or not *scrP* is expressed by this plasmid.

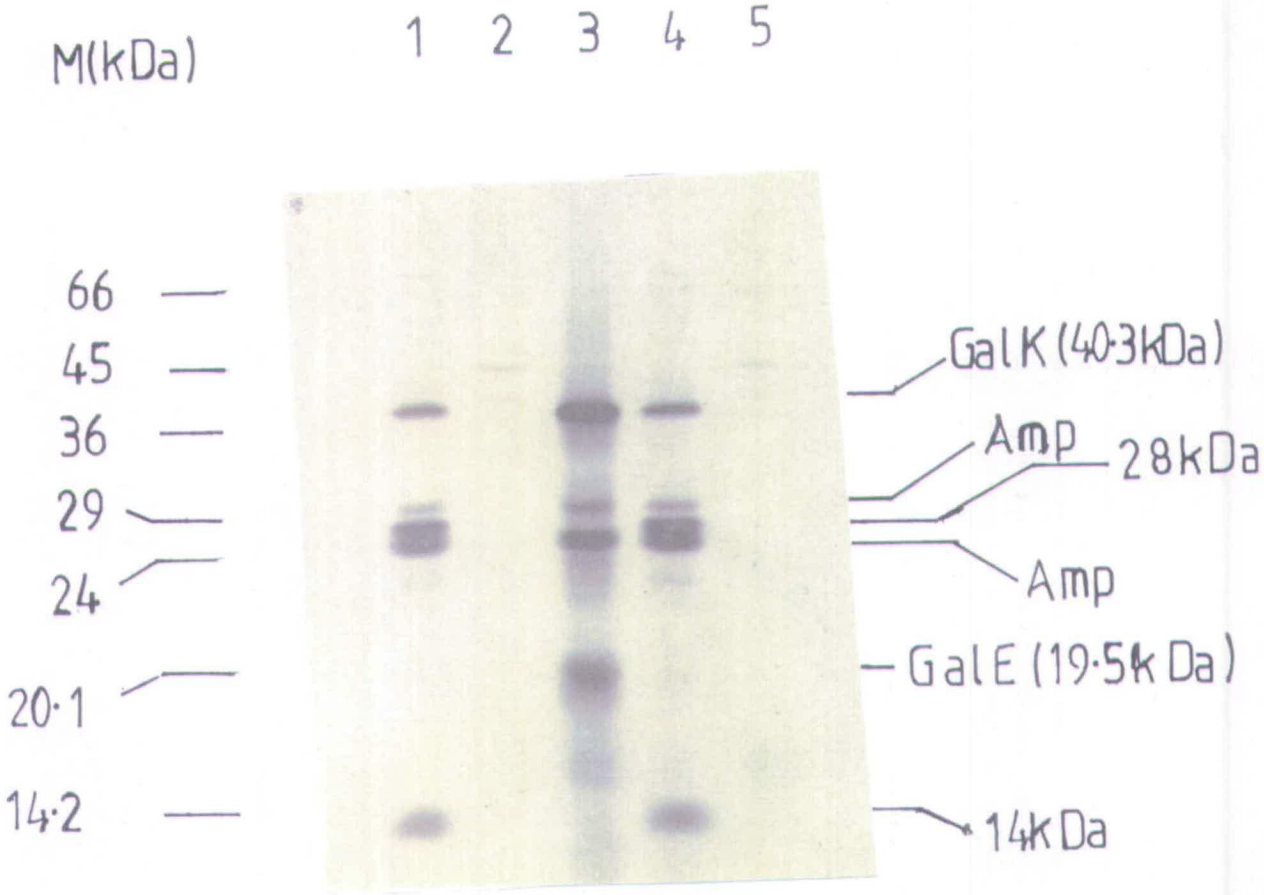
Non-transformed minicells and minicells harbouring pFMT1 or its parent plasmid, pJW11 were separated from P678.54 and its derivatives and labelled with [^{35}S]Met as described in Sections 2.9.1 and 2.9.2. The labelled proteins in crude minicell extracts were then separated by SDS-PAGE and observed by autoradiography as described in Sections 2.10 and 2.2.6.

As figure 4.8 shows, polypeptides of estimated MW ~14 kDa and 28 kDa were synthesised **solely** by minicells harbouring pFMT1, and presumably originiate from sequences within the *Hin*dIII-*Eco*RI insert. The plasmid also expressed *galK* and AmpR, as did pJW11, which in addition expressed the *galE'* gene product. In pFMT1 *galK* expression is almost certainly dependent on promoter sequences within the cloned insert, as the *gal* promoter and the *galE'* fragment had been deleted during the construction of the plasmid. The significantly lower level of GalK synthesis from pFMT1, in comparison with pJW11 presumably reflects the relative strengths of the *gal* promoter and the promoter(s) within the cloned insert in pFMT1 which are responsible for *galK* transcription.

Figure 4.8

Proteins encoded by pFMT1 in minicells. Autoradiography of a 10% polyacrylamide gel. Tracks 1-5 show gene products in minicell extracts of E. coli strain P678.54 bearing plasmids: (1 and 4) pFMT1; (3) pJW11 and (2 and 5) "no plasmid". Proteins encoded by pFMT1 and/or pJW11 are marked to the right of the photograph and are identified and/or their sizes given. The sizes and positions of molecular mass standards are given to the left of the photographs. Amp denotes beta-lactamase (both the nascent and mature proteins are indicated). GalK denotes galactokinase and GalE denotes C-terminally truncated UDP-galactose 4-epimerase. This experiment was performed by Fiona M. Townsley.

FIGURE 4.8


M(kDa)
1
2
3
4
5
66
45
36
29
24
20·1
14·2
GalK (40·3kDa)
Amp
28kDa
Amp
GalE (19·5k Da)
14k Da

By consulting the physical map of the *E. coli* K12 chromosome (Kohara *et al* 1991, Rudd *et al* 1992) it was ascertained that the only known gene likely to be located within the cloned insert of pFMT1 is part of *arcB*. (The *rpoN* gene and 3 associated open reading frames map outside the insert, to the left of the *Hin*dIII site). However the 5' terminus of the *arcB* gene, including its ATG start codon and upstream transcription and translation elements, are missing from the pFMT1 insert. Not surprisingly ArcB (MW 88 kDa) was not produced in the minicells. Therefore the 28 kDa and 14 kDa polypeptides may be the products of hitherto unmapped genes lying downstream of ArcB, and one or both might be transcribed from the promoter which is driving *galK* transcription in pFMT1.

In attempts to immunoprecipitate proteins from the minicell extracts with the anti-2.2-peptide antibody used in the original identification of SCRP27-A, the ~28 kDa polypeptide was not significantly recovered. Whilst this might reflect a difference in the accessibility of the relevant epitopes on SCRP-27A protein in solution as opposed to its immobilised state on Western blotting membranes, it is equally possible that the 28 kDa polypeptide produced by pFMT1 is not SCRP-27A. If so, *scrP* may not be expressed detectably in pFMT1.

4.6.1 Nucleotide Sequence of *scrP*27A

Initial nucleotide sequence data were generated using

*scrp*27A-1 as a primer for sequencing mp18-15 RF DNA, as described in Section 2.7. Further nucleotide sequence information was generated using both single and RF DNAs of mp18-15 and mp19-23 as templates, with synthetic oligonucleotide primers designed to be complementary to already determined *scrP* sequence. The sequencing strategy is presented in figure 4.9.

910 bp of sequence were determined in both DNA strands (Figure 4.10). One long open reading frame was apparent within the nucleotide sequence, extending for 218 codons (excluding the TAA stop codon) between positions 145 and 798 inclusive. A sequence which closely resembles the bacterial Shine and Dalgarno consensus signal is located shortly upstream of the ATG start codon of this open reading frame, as are -10 and -35 elements of a putative sigma 70 dependent promoter. Comparison of the predicted polypeptide sequence encoded by the open reading frame with the N-terminal aa sequence determined for SCRP-27A, (underlined in Figure 4.10) confirmed the identity of the open reading frame as the *scrP* gene and verified that the probable translation start point is that indicated above.

4.6.2 Codon Usage in *scrP*

The codon usage of *scrP* was determined using the codon frequency programme run on the GCG package (Table 4.1).

The Fop value of 0.61 shows that there is no strong bias in the coding sequence towards codons recognised by major tRNA species of *E. coli*. It is characteristic of moderately to weakly expressed *E. coli* genes.

Table 4.1 Codon Usage in *scrP*

First Letter	Second Letter U	C	A	G
U	UUU Phe 5	UCU Ser 1	UAU Tyr 2	UGU Cys 2
	UUC Phe 3	UCC Ser 0	**UAC** Tyr 0	UGC Cys 3
	UUA Leu 4	UCA Ser 1	UAA Ochre 1	UGA Opal 0
	UUG Leu 6	UCG Ser 1	UAG Amber 0	UGG Trp 3
C	CUU Leu 2	CCU Pro 2	CAU His 4	**CGU** Arg 3
	CUC Leu 0	CCC Pro 1	CAC His 0	**CGC** Arg 4
	CUA Leu 0	CCA Pro 2	CAA Gln 1	CGA Arg 2
	CUG Leu 10	**CCG** Pro 7	**CAG** Gln 6	CGG Arg 1
A	AUU Ile 9	**ACU** Thr 2	AAU Asn 4	AGU Ser 1
	AUC Ile 5	**ACC** Thr 5	AAC Asn 1	AGC Ser 5
	AUA Ile 2	ACA Thr 1	AAA Lys 3	AGA Arg 1
	AUG Met 7	ACG Thr 4	AAG Lys 6	AGG Arg 0
G	**GUU** Val 4	**GCU** Ala 2	GAU Asp 10	**GGU** Gly 6
	GUC Val 4	GCC Ala 6	GAC Asp 3	GGC Gly 5
	GUA Val 1	**GCA** Ala 6	**GAA** Glu 12	GGA Gly 2
	GUG Val 6	**GCG** Ala 14	GAG Glu 2	GGG Gly 2

Fop = 0.61

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

4.6.3 Nucleotide Sequence Comparisons

Comparison of the *scrP* nucleotide sequence with the EMBL database revealed significant similarity to only one DNA sequence; the 5′ 89 nt of the sequence upstream of *scrP* overlap the 3′ terminus of the known sequence of the *E. coli arcB* locus (Iuchi *et al* 1990). The matching sequences were found to be identical except for a single substitution of G for T at position 71 within the sequence determined here (nt 2516 in the published sequence). The substitution occurs within the 250 bp intergenic region between *arcB* and *scrP*, which would be transcribed in the same direction counter-clockwise on the *E. coli* chromosome.

4.7 Transcriptional Analysis of the *scrP* Gene

Analysis of the *arcB-scrP* intergenic region shortly upstream of *scrP* revealed possible -10(TGTGAT) and -35(TTGTTA) elements, separated by 17 bp, which might allow sigma 70 dependent initiation of *scrP* transcription. On the other hand the relative orientation and close proximity of *scrP* to arcB raise the possibility that these genes may be co-transcribed. Therefore an attempt was made to map the 5′ termini of *scrP* mRNAs by primer extension analysis to ascertain whether the putative *scrP* promoter is functional.

Figure 4.9

Strategy for sequencing *scr*P. The top part shows the restriction map of the *E. coli* genomic DNA fragment cloned in mp18-15 and mp19-23, which are shown below the map. Small arrows indicate the direction and approximate extent of sequence reactions primed with different oligonucleotides. Only the relevant restriction sites are shown. The following abbreviations are used for restriction enzymes: H, *Hin*dIII; R, *Eco*RI and V, *Pvu*II.

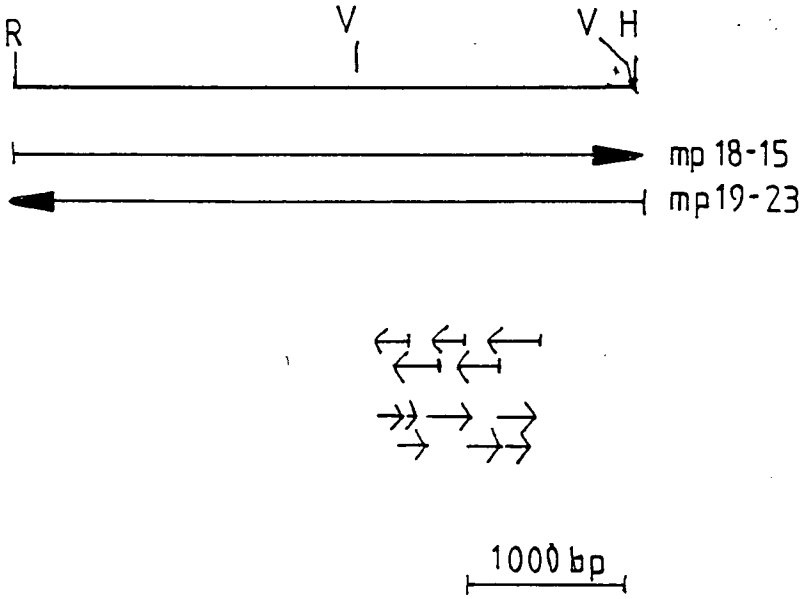
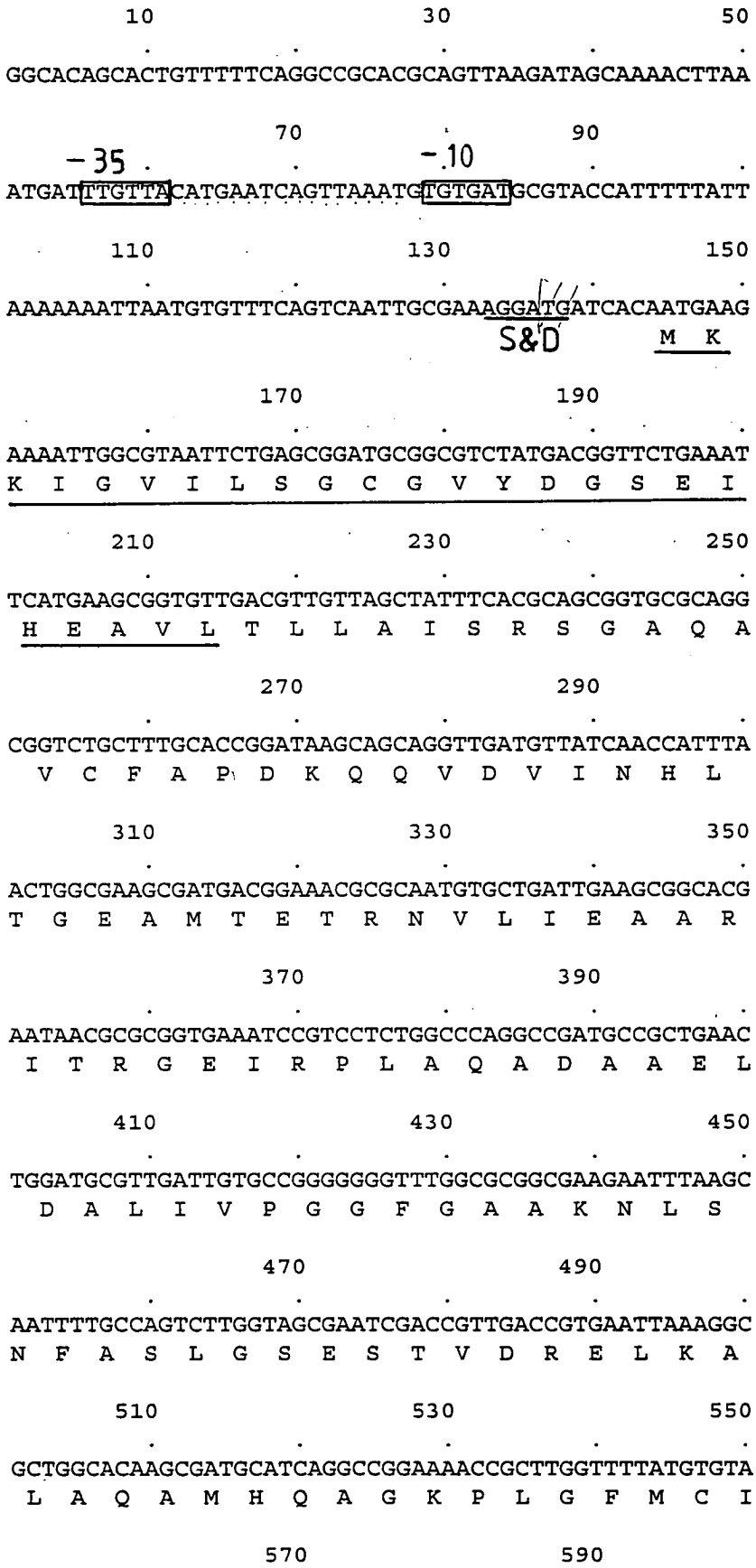


FIGURE 4.9

Figure 4.10

Nucleotide sequence of the *scrP* gene and predicted aa sequence of the encoded protein. The nucleotide sequence of the *scrP* gene and its flanking regions is shown in the 5′ to 3′ direction. The predicted aa sequence of the SCRP-27A protein is shown below the relevant codons. The underlined aa sequence is identical to the N-terminal aa sequence reported by Ueshima *et al* (1992) except that aa 11 (C in the predicted sequence) could not be identified by aa sequencing (*ibid*). The likely Shine and Dalgarno sequence and the -10 and -35 elements of a putative promoter upstream of *scrP* are underlined and boxed respectively. The sequence was submitted to DDJB and was given the accession number D13188.


10
30
50
GGCACAGCACTGTTTTTCAGGCCGCACGCAGTTAAGATAGCAAAACTTAA
70
90
-35
-10
ATGATTTGTTACATGAATCAGTTAAATGTGTGATGCGTACCATTTTTATT
110
130
150
AAAAAAATTAATGTGTTTCAGTCAATTGCGAAAGGATGATCACAATGAAG
S&D
M K
170
190
AAAATTGGCGTAATTCTGAGCGGATGCGGCGTCTATGACGGTTCTGAAAT
K I G V I L S G C G V Y D G S E I
210
230
250
TCATGAAGCGGTGTTGACGTTGTTAGCTATTTCACGCAGCGGTGCGCAGG
H E A V L T L L A I S R S G A Q A
270
290
CGGTCTGCTTTGCACCGGATAAGCAGCAGGTTGATGTTATCAACCATTTA
V C F A P D K Q Q V D V I N H L
310
330
350
ACTGGCGAAGCGATGACGGAAACGCGCAATGTGCTGATTGAAGCGGCACG
T G E A M T E T R N V L I E A A R
370
390
AATAACGCGCGGTGAAATCCGTCCTCTGGCCCAGGCCGATGCCGCTGAAC
I T R G E I R P L A Q A D A A E L
410
430
450
TGGATGCGTTGATTGTGCCGGGGGGGTTTGGCGCGGCGAAGAATTTAAGC
D A L I V P G G F G A A K N L S
470
490
AATTTTGCCAGTCTTGGTAGCGAATCGACCGTTGACCGTGAATTAAAGGC
N F A S L G S E S T V D R E L K A
510
530
550
GCTGGCACAAGCGATGCATCAGGCCGGAAAACCGCTTGGTTTTATGTGTA
L A Q A M H Q A G K P L G F M C I
570
590

```
                                                                    .
TTGCCCCGGCGATGCTGCCGAAAATTTTCGATTTCCCGCTGCGTTTGACC
  A  P  A  M  L  P  K  I  F  D  F  P  L  R  L  T

       610                   630                   650
        .                     .                     .
ATCGGTACTGATATCGATACCGCAGAAGTGCTGGAAGAGATGGGCGCGGA
I  G  T  D  I  D  T  A  E  V  L  E  E  M  G  A  E

                  670                   690
                   .                     .
GCATGTGCCGTGTCCTGTCGATGATATCGTGGTTGATGAAGACAATAAGA
 H  V  P  C  P  V  D  D  I  V  V  D  E  D  N  K  I

       710                   730                   750
        .                     .                     .
TTGTCACCACCCCAGCATATATGCTGCGCAGAACATTGCAGAAGCGGCGA
  V  T  T  P  A  Y  M  L  R  R  T  L  Q  K  R  R

                  770                   790
                   .                     .
GCGGCATTGATAAGCTGGTTTCCCGCGTGCTGGTTCTGGCTGAATGAGTA
A  A  L  I  S  W  F  P  A  C  W  F  W  L  N  E  *

       810                   830                   850
        .                     .                     .
AAAGCCGCTTAACGGTGTTTAGTTTCGTTCGCCGTTTTCTTTTGCGGTTA

                  870                   890
                   .                     .
ATGGTTGTCCTCGCCGTTTTCTGGGGCGGGGGCATCGCGTTGTTTAGCGT

       910
        .
TGCCCTGTTC
```

4.7.1 Primer Extension Analysis of *scrP* mRNA

Total *E. coli* RNA was isolated and subjected to primer extension analysis as described in Sections 2.5.7 and 2.8.7 using the deoxyribonucleotide 876L[5'-CAACGTCAACACCGCTTC-3'] as a primer. (876L is complementary to nt 222 through 206 of the *scrP* coding strand). Figure 4.11 shows that a large number of reverse transcripts were generated which terminated on the mRNA templates at several locations within the *arcB-scrP* intergenic region.

A pair of relatively minor reverse transcripts did terminate near the likely start point of the putative *scrP* promoter. These transcripts correspond to nucleotides C91 and A92, located 7 and 8 bp respectively downstream of the -10 element of the putative promoter. However the majority of reverse transcripts terminated well beyond this region, strongly suggesting that most *scrP* transcription initiates far upstream - probably due to co-transcription with *arcB* - although a low level of independent transcription from positions 91/92 is not excluded.

Whilst the large number of reverse transcripts is suggestive of premature termination within the mRNA templates, possibly at sites of strong secondary structure, it is possible that some of the more prominent

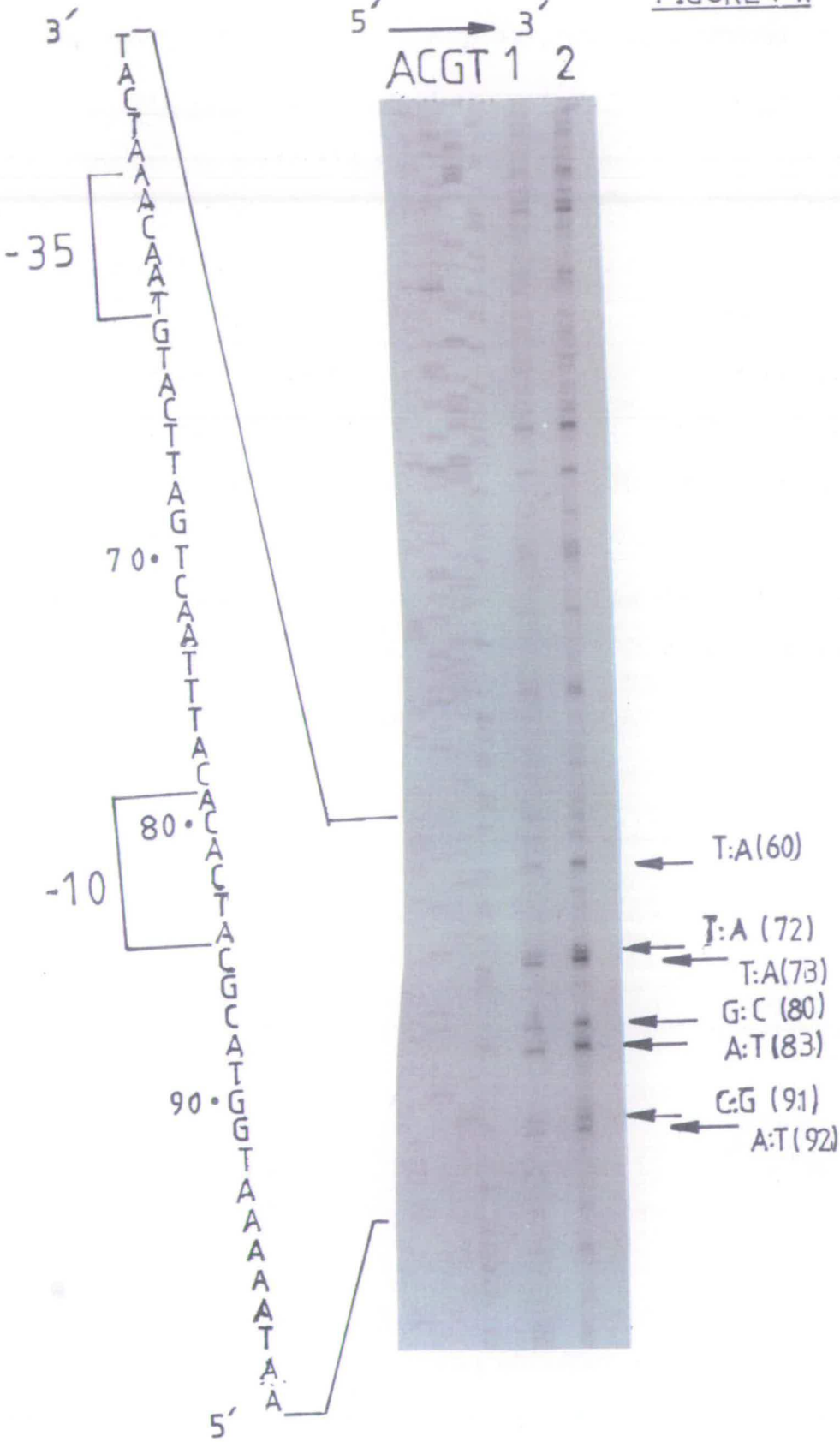
termini may be of functional significance. In particular 5 more prominent transcripts terminated shortly upstream of the *scrP* coding region. These termination sites - T60, T72, T73, G80 and A83 - could represent mRNA processing sites, rather than merely more frequent premature termination by reverse transcriptase.

Confirmation of the proposed *arcB-scrP* co-transcription will require further more detailed transcription analysis.

Figure 4.11

Primer extension analysis of *scrP* transcripts. Oligonucleotide primer 876L was hybridised with 30 (track 1) or 60 (track 2)µg respectively of RNA extracted from *E. coli* strain W3110 and extended with reverse transcriptase. The hybridisation temperature was 40°C. The cDNA products were analysed on a 6% polyacrylamide gel. A DNA sequence ladder, primed with the same oligonucleotide, using M13 mp18-15 ssDNA as template, served as marker. The DNA sequence (bottom strand) of the *scrP* 5' flanking region is shown to the left of the photograph; the -10 and -35 sequences are indicated. Arrows point to prominent cDNA bands described in the text; the corresponding base pairs in the DNA sequence are indicated and their positions within the sequence given.

FIGURE 4.11



4.8 SCRP-27A Protein

The aa composition of the predicted SCRP-27A protein is shown in Table 4.2. The calculated MW of 23,647 is slightly less than the 27000 value estimated by SDS-PAGE analysis (Ueshima *et al* 1992). Hydrophilicity analysis (figure 4.12) revealed no major regions of hydrophobic or charged amino acids within the SCRP-27A polypeptide sequence. However, two short stretches of amino acids near the N-terminus of the polypeptide sequence, which are moderately hydrophobic, might allow SCRP-27A to associate with the cell membrane.

Comparison of the SCRP-27A aa sequence with the NBRF-PIR SWISS-PROT and translated EMBL databases, revealed no significant similarity between SCRP27-A and known polypeptide sequences. Similar comparison with a database of known functional and/or structural aa sequence motifs also proved negative. Failure to detect significant similarity between SCRP-27A and known bacterial RNA polymerase sigma subunit aa sequences suggests that *scrP* is unlikely to represent a new sigma gene. This is consistent with the observation that SCRP-27A does not associate with RNA polymerase core *in vitro* (Ueshima *et al* 1992).

Table 4.2 Amino Acid Composition of SCRP-27A[a]

Amino Acid	Number of Residues Per Subunit	Mol %
D	13	5.963
N	6	2.752
T	12	5.505
S	9	4.128
E	15	6.881
Q	7	3.211
P	11	5.064
G	15	6.881
A	28	12.884
C	5	2.994
V	15	6.881
M	7	3.211
I	16	7.339
L	22	10.092
Y	2	0.917
F	8	3.670
K	9	4.128
H	4	1.835
R	11	5.046
W	3	1.376

MW = 23647.09 Isoelectric Point = 4.89

[a] inferred from the DNA sequence of the *scrP* gene.

Figure 4.12

Hydrophilicity plot for the predicted aa sequence of SCRP-27A. The hydrophilicity value of each aa residue is plotted against its position in the polypeptide sequence, starting with the amino terminus (Kyte and Doolittle 1982). The aa scale is shown above the plot.

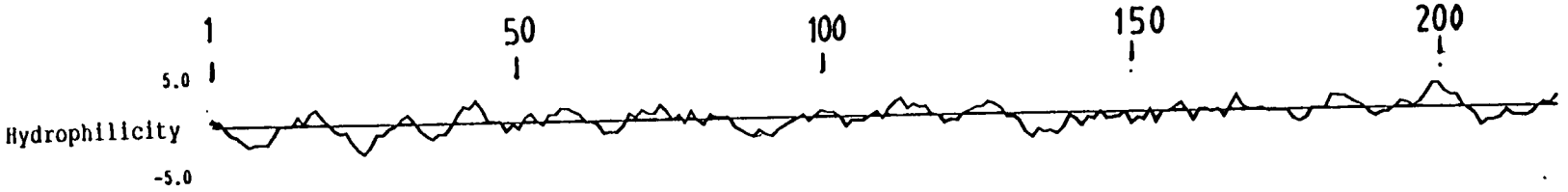


FIGURE 4.12

4.9 Characterisation of *scrP* by Gene Disruption

Database comparison revealed no significant similarity between SCRP-27A and known protein sequences (see Section 4.8), and so provided no clues as to its function. Therefore the genomic copy of *scrP* was replaced with a mutated copy (disrupted *in vitro* by insertion of a *kan* cassette) in order to investigate the cellular function of the SCRP-27A protein.

4.9.1 Disruption of *scrP in vitro*

The *scrP* gene was first disrupted *in vitro* by insertion of a kanamycin resistance cassette into its coding sequence. Inspection of the *arcB* and *scrP* DNA sequences revealed a single *NsiI* restriction site (at bp 518) within the *scrP* coding sequence which was subsequently verified as the sole *NsiI* target within pFMT1 by restriction analysis (data not shown). The ~1.3kb *PstI* fragment of pUC4-KISS, which carries the Tn*901* kanamycin resistance gene was then inserted into the *NsiI* site of pFMT1 to produce pDSA28 (figure 4.13).

4.9.2 Transfer of the disrupted *scrP* on to the *E. coli* Chromosome

A transducing phage line was generated by preparing plate lysates of lambda 523 on *E. coli* TG1 which had been transformed with pDSA28, as described in Section 2.4.5. The lysate was then titrated and used to infect the RecD

strain NM621 at a MOI of 1, as described in Section 2.4.8. Transductants were recovered on LB plates containing kanamycin 50μgml^{-1}). 14/14 transductants recovered were KanR and AmpS. The latter evidence that the transductants were free from pDSA28 was confirmed by small scale plasmid isolation (data not shown).

The *scrP* disruption was verified by Southern hybridisation analysis of genomic DNA isolated from 3 transductant strains using as probe the 1.7kb *Pvu*II fragment of pFMT1-[^{32}P] labelled by random priming), as described in Section 2.8.1.

As figure 4.14 shows, the *Pvu*II probe hybridised with a single ~2.3kb *Pvu*II fragment of NM621 DNA whereas the size of the *Pvu*II fragments in all three transductant strains is ~0.8kb larger, at ~3.1kb. Correspondingly, the probe identified a single 6.7kb *Hind*III fragment of NM621 DNA whereas two fragments of ~2.0kb and 5.5kb were identified in all three transductants (tracks 6-8) due to the situation of a second *Hind*III site within the KanR cassette. These results confirm that the *scrP* gene has been replaced on the *E. coli* K12 chromosome by the disrupted allele in all three transductant strains and indicate that *scrP* is inessential for the growth and viability of *E. coli* cells under the conditions employed. One of the transductants was selected for further characterisation and named DSA100.

Figure 4.13

Summary of the construction of pDSA28. Open rectangles represent the relevant genes; arrows indicate the transcriptional orientations. Hatched boxes represent the ~1.3 kilobase *kan* cassette derived from pUC4-KISS. *amp* encodes beta-lactamase; *gal*K encodes galactokinase and KanR encodes kanamycin resistance. Only the relevant restriction sites are shown. The following abbreviations were used for restriction enzymes: B, *Bam*HI; H, *Hind*III; N, *Nsi*I; P, *Pst*I; R, *Eco*RI and V, *Pvu*II. N/P* denotes that *Nsi*I and *Pst*I "ends" were ligated together thus preventing regeneration of either site.

FIGURE 4J3

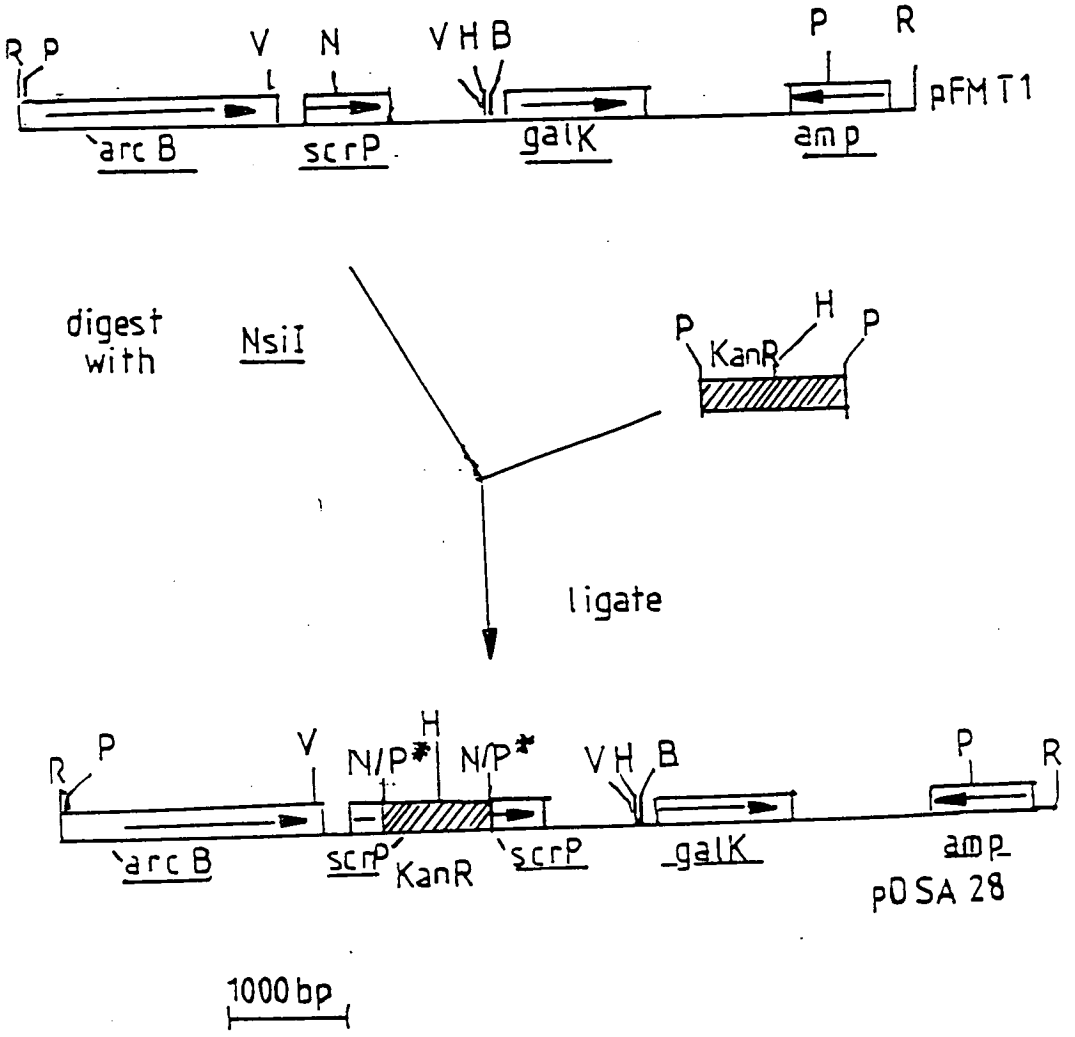

R P
V
N
V H B
P
R
pFMT1
'arc B
scrP
galK
amp
digest
with
NsiI
P
KanR
H
P
ligate
R
P
V
H
N/P*
N/P*
V H B
P
R
'arc B
scrP
KanR
'scrP
galK
amp
pOSA 28
1000 bp

Figure 4.14

Verification of transfer of *scr*P :: KanR onto the E. coli K12 chromosome by Southern hybridisation analysis. Hybridisation and washing were carried out as described in section 2.8.4. Tracks 1 and 5 show *E. coli* strain NM621 chromosomal DNA and tracks 2, 3, 4, 6, 7 and 8 show transductant DNAs. The sizes of marker frgments are given (in base pairs) to the right of the photograph. The following abbreviations are used for restriction enzymes: H, *Hin*dIII and V, *Pvu*II.

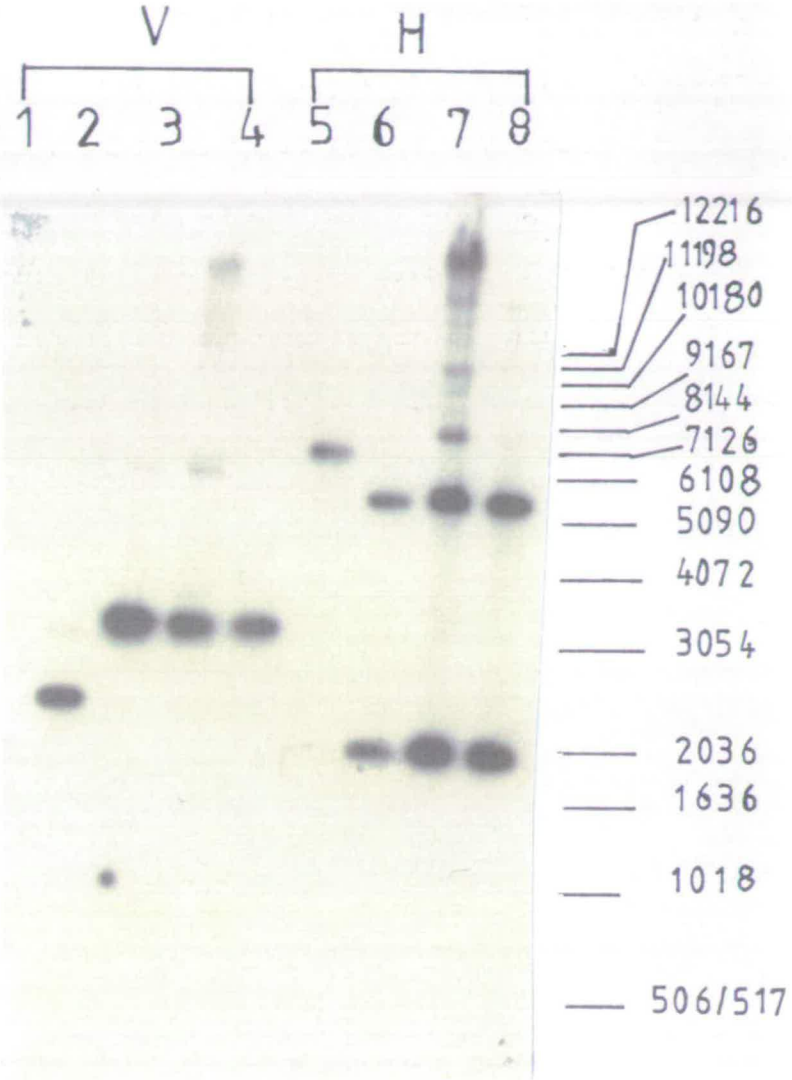

V
H
1 2 3 4 5 6 7 8
12216
11198
10180
9167
8144
7126
6108
5090
4072
3054
2036
1636
1018
506/517

FIGURE 4.14

Two features of Figure 4.14 require further comment. The estimated sizes of the probe positive *Hin*dIII (~6.7kb and *Pvu*II (~2.3kb) fragments of NM621 DNA are approximately 1.7 and 0.6kb respectively larger than shown on the *E. coli* K12 physical map (see Figure 4.4 These differences might be explained if the *Hin*dIII site near 3420kb and the *Pvu*II site near 3416kb on the physical map are not present on the NM621 chromosome and presumably reflects small differences within the DNA sequences of NM621 and W3110 (the source of chromosomal DNA used in the construction of the Kohara library).

Secondly, the size of the probe positive *Pvu*II fragment in <u>all 3</u> transductant DNAs is only 3.1 kb. This is 0.5 kb less than the 3.6 kb figure expected from the insertion of the 1.3 kb *kan* cassette into the 2.3 kb *Pvu*II fragment of NM621 DNA. This may have resulted from deletion of DNA from the *arcB-scrP* region during an early stage in the transduction of the plasmid borne *scrP::kan* allele onto the NM621 chromosome. The site of the proposed deletion was defined by examination of the sizes of the probe positive *Hin*dIII fragments in the transductant DNAs. The expected sizes of these fragments were ~6.0 kb and ~2.0 kb. The actual observed sizes estimated above clearly showed that the deletion had occurred in the larger of the *Hin*dIII fragments which carries *arcB* and *scrP* DNA upstream of the *Nsi*I site at 518 bp in the *scrP* DNA sequence. Thus it was concluded

that ~0.5 kb of DNA had been lost from the transductant chromosome, somewhere between the *Pvu*II site at position 2199 in the published *arcB* sequence (Iuchi and Lin 1990) and the *Hin*dIII site within the *kan* cassette. The fact that the transductants were Kan^r and $ArcB^+$ (see Section 4.9.5 below) further suggests that the deletion occurred between the 3' terminus of *arcB* and the *Nsi*I site in *scrP* (bp2963 on the *arcB* scale). Obviously further verification of this is necessary to ensure that the 3' terminus of *arcB* has not been affected by the proposed deletion.

4.9.3 Construction of DSA101

DSA100 is *recD* and therefore unsuitable for maintenance of colE1 type replicons (such as pFMT1). Since characterisation of the *scrP* :: *kan* disruption might be aided by future complementation analysis using pFMT1, the *scrP* :: *kan* disruption was moved to a $RecD^+$ background by transduction with P1vir.

A P1vir plate lysate was prepared on DSA100 and used to infect MM38 (*arg*G6) as described in Sections 2.4.2 and 2.4.3. *argG*, which is required for arginine biosynthesis, is located near 69 minutes on the *E. coli* genetic map (Bachman 1990) and was used to verify *scrP* disruption by the simultaneous transduction of MM38 to Arg^+ and Kan^r. Arg^+ transductants were selected on Spizizen glucose minimal agar, then patched onto LB agar

containing kanamycin ($50\mu g ml^{-1}$). 28% of Arg^+ transductants were also kanamycin resistant which (using Wu's formula (1966) estimates the map distance between *argG* and *scrP :: kan* as 0.69 minutes on the MM38 chromosome. This is consistent with the physical map location of *scrP* near 3416 kb (69.75 minutes) [see Section 4.2.2]. MM38 *scrP :: kan* was renamed DSA101.

4.9.4 Characterisation of DSA100 and DSA101

The ability to disrupt the *scrP* gene *in vivo* by insertion of the *Tn901* kanamycin resistance gene into its coding sequence, demonstrated that it is not essential for *E. coli* viability either in rich medium (LB) or nutrient limited medium (glucose minimal medium). In order to further assess the effects of the *scrP :: kan* disruption on *E. coli* growth, the growth rates of DSA100 and DSA101 were compared with those of their parent strains NM621 and MM38 in rich medium, and of DSA100 with NM621 in Spizizen glucose minimal medium. As figure 4.15 shows there was no significant difference between the doubling times of strains carrying the *scrP :: kan* disruption and the $scrP^+$ strains when grown in LB or nutrient broth. Growth in minimal medium was similarly unimpaired by the *scrP :: kan* genotype (Figure 4.16), suggesting that the *scrP* disruption does not significantly affect major macromolecular synthesis or other essential pathways.

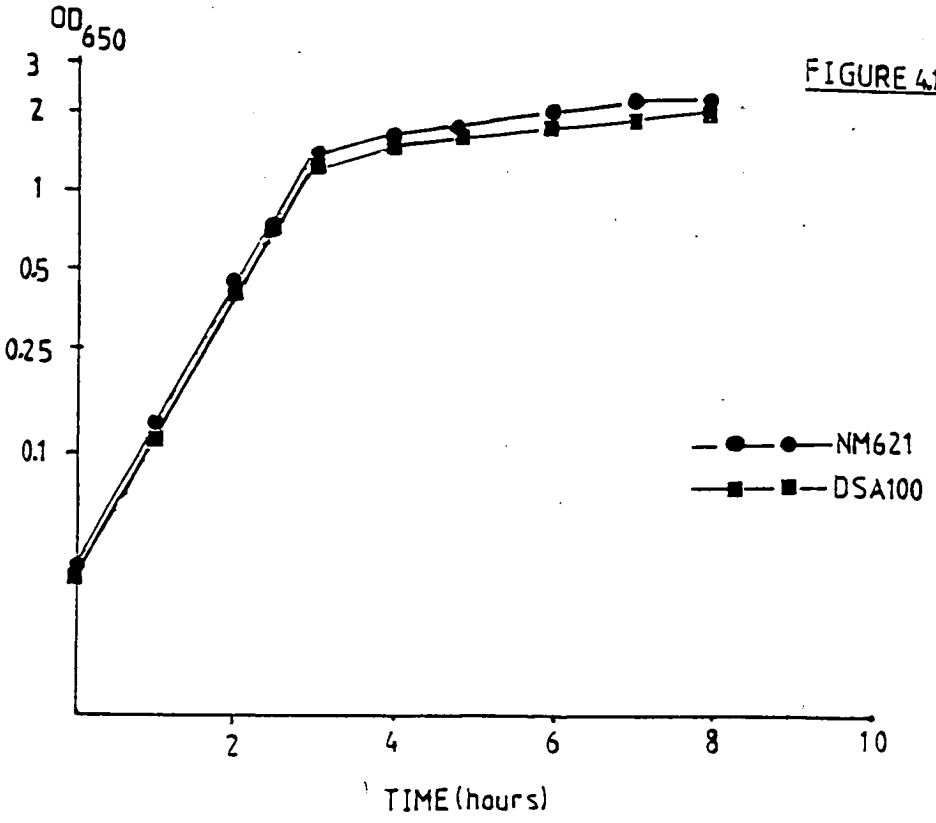
The ability of DSA100 and DSA101 to grow on a variety of carbon sources and at various temperatures between 30 and 46°C was also assessed. The ability of both strains to grow anaerobically at 37°C was also tested. The results are summarised in Table 4.3 and clearly show that SCRP-27A is not required for growth under the conditions employed.

One particularly relevant result in Table 4.3 is the ability of DSA100 and DSA101 to grow on minimal medium containing galactitol as the sole carbon source. The *kba* gene, encoding a temperature sensitive ketose-bis-phosphate aldolase has been genetically mapped near 69.5 minutes on the *E. coli* K12 chromosome (Lengler 1977). Kba is required for galactitol metabolism between 30°C and 37°C, but its thermal inactivation prevents growth at temperatures exceeding 42°C. The ability of DSA100 and DSA101 to grow on galactitol plates at 30°C clearly suggests that SCRP-27A is **not** encoded by *kba*.

Figure 4.15

The growth rates of *E. coli* strains DSA100 and DSA101 and their *scrp*$^+$ parents in rich medium (LB). O.D. 650; optical density at 650nm. Where appropriate kanamycin was added to the medium at a final concentration of 50μg ml^{-1}.

FIGURE 415


OD650
3
2
1
0.5
0.25
0.1
2
4
6
8
10
TIME (hours)
NM621
DSA100

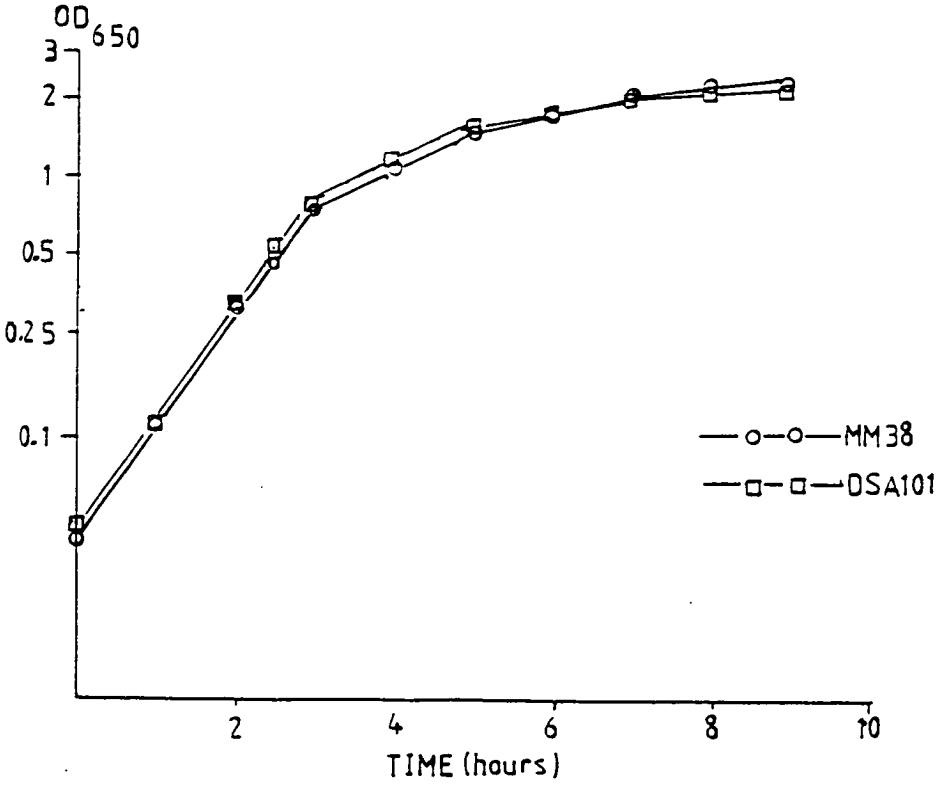

OD650
3
2
1
0.5
0.25
0.1
2
4
6
8
10
TIME (hours)
MM38
DSA101

Figure 4.16

The growth rates of *E. coli* strains DSA100 and NM621 in nutrient limited medium (Spizizen minimal medium supplemented with glucose, thiamine, proline and leucine at the concentrations stated in section 2.1.3). Kanamycin was included in the medium at 30μg ml^{-1} to culture DSA100. O.D. 650; optical density at 650nm.

FIGURE 4.16

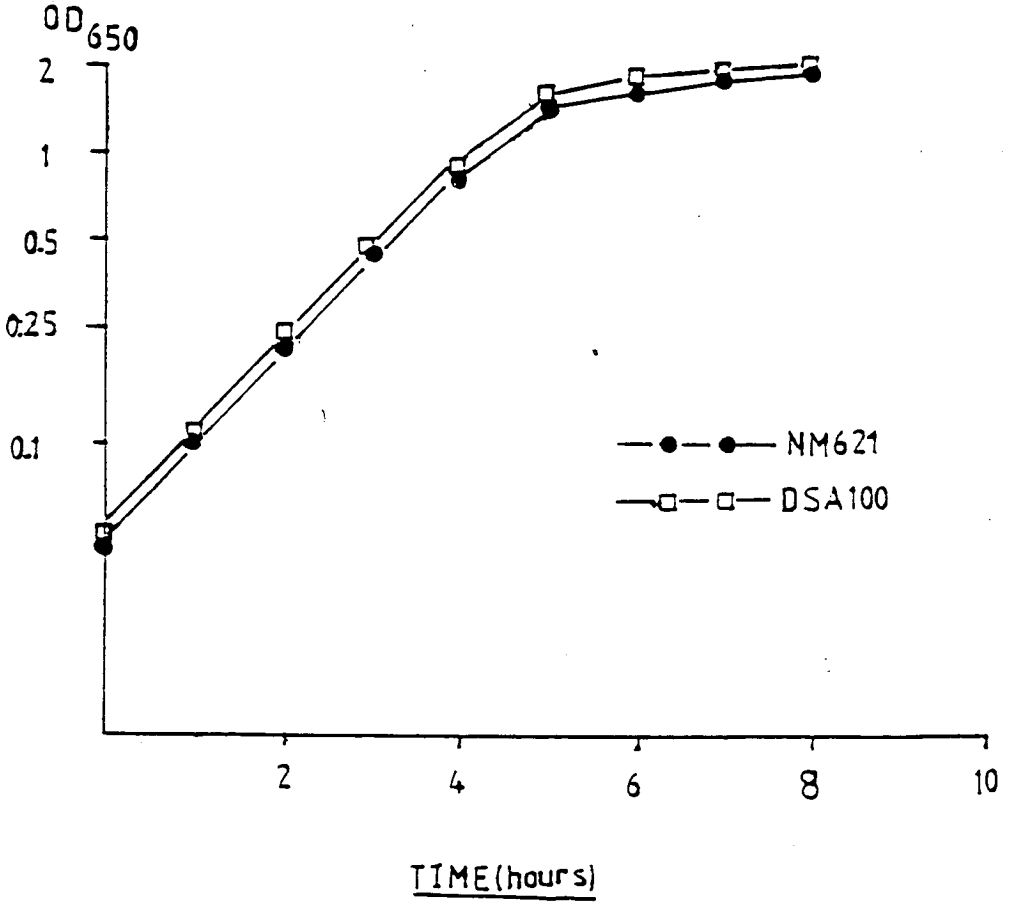


TABLE 4.3 **GROWTH OF *scrP :: kan* STRAINS UNDER VARIOUS CONDITIONS**

Strain	Medium	+ O_2				- O_2
		30°C	37°C	42°C	46°C	37°C
NM621	LB	+	+	+	+	+
	Glucose	+	+	+	ND	+
	Glycerol	+	+	+	ND	ND
	Xylose	+	+	ND	ND	ND
	Maltose	+	+	ND	ND	ND
	Galactitol	+	ND	-	-	ND
DSA100	LB	+	+	+	+	+
	Glucose	+	+	+	ND	+
	Glycerol	+	+	+	ND	ND
	Xylose	+	+	ND	ND	ND
	Maltose	+	+	ND	ND	ND
	Galactitol	+	ND	-	-	ND
MM38	LB	+	+	+	+	+
	Glucose	+	+	+	ND	+
	Glycerol	+	+	+	ND	ND
	Galactitol	+	ND	-	-	ND
DSA101	LB	+	+	+	+	+
	Glucose	+	+	+	ND	+
	Glycerol	+	+	+	ND	ND
	Galactitol	+	ND	-	-	ND

0.5ml of an overnight culture of each strain was washed 3 times and resuspended in 0.5ml of bacterial buffer. Aliquots were then streaked onto agar media and plates incubated for between 14 and 72 hours. Where appropriate kanamycin was added to media at a final concentration of 50 $ugml^{-1}$. Sugars, amino acids and thiamine were added to Spizizen minimal media at the concentrations stated in Section 2.1.3. Strains were grown anaerobically for 18-72 hours in Oxoid anaerobic gas jars in an environment of H_2 and CO_2 using Oxoid gas packs according to the manufacturer's instructions.

Symbols: **+ denotes growth, - denotes no growth, ND denotes no data. $+O_2$ and $-O_2$ denote aerobic and anaerobic growth conditions respectively.**

4.9.5 Dye Sensitivity Tests

The transcription data presented in Section 4.7.1 tentatively suggest that *scrP* may be co-transcribed with the *arcB* gene. ArcB is a transmembrane sensor regulator protein involved in the regulation of metabolic pathways in response to oxygen availability (Iuchi *et al* 1990). Under anaerobic conditions ArcB activates a transcription regulator - ArcA, encoded near 0 minutes on the *E. coli* chromosome - which represses transcription of several genes involved in aerobic metabolism, including citric acid cycle genes, fatty acid degradation pathways and glyoxylate cycle genes. The ArcAB system is also thought to positively affect the transcription of *cyd* (encoding cytochrome D-oxidase) which is expressed during stationary phase and anaerobiosis (reviewed by Spiro and Guest 1991).

ArcA is separately involved in the regulation of F-pilus expression and other conjugal fertility functions in conjunction with a second transmembrane sensor protein - CpxA. In both systems the activation of ArcA is brought about through its phosphorylation by one or other of ArcB or CpxA (quite possibly at different sites). Mutations which inactivate either ArcA or ArcB result in elevated anaerobic expression of aerobic respiratory functions. For example an *sdh-lacZ* transcriptional fusion is expressed ~17 fold more

strongly in *arcA* or *arcB* mutants than in otherwise isogenic control strains (Iuchi and Lin 1989 and 1990). Arc mutations also result in sensitivity of *E. coli* cells to redox dyes such as methylene blue and its analogue toluidine blue (Buxton *et al* 1983, Iuchi and Lin 1989). The latter phenotype provides an easy means of testing the effects of mutations on the regulation of aerobic respiratory control. Therefore it was decided to test the sensitivity of DSA100 and DSA101 to toluidine blue and methylene blue as a preliminary investigation of the possible involvement of *scrP* in Arc function, suggested by the putative co-transcription of *arcB* and *scrP*.

Dye sensitivity tests were carried out as described in Section 2.3.3. As Table 4.4 shows, the growth of DSA100 and DSA101 was unaffected by either toluidine blue or methylene blue, whereas the *arcA*1 strain ECL585 failed to grow on medium containing either dye.

An unexpected result was the apparent dye sensitivity of MM38, the parent of DSA101, suggesting that this strain carries a previously unidentified mutation in either *arcA* or *arcB*. If this is so, it is likely that the mutation is in *arcB* as DSA101 (dye resistant) was constructed by co-transduction of MM38 to Arg^+ and *scrP :: kan*. It is therefore possible that MM38 was also simultaneously transduced to $ArcB^+$ during construction of DSA101.

Obviously it would be desirable to verify this by, for example, complementation analysis or marker rescue using the cloned *arcB* gene. The observed dye resistance of DSA100 and DSA101 does not exclude the possibility that SCRP-27A participates in the functioning of the Arc regulatory system. To investigate this further, the effects of the *scrP :: kan* insertion on the expression of an *sdh-lacZ* fusion are underway, whilst possible further investigations are discussed in Section 4.10 below.

TABLE 4.4 DYE SENSITIVITY TESTS

	Dye [a]	
Strain	**Toluidine Blue**	**Methylene Blue**
NM621	R	R
DSA100	R	R
MM38	S	S
DSA101	R	R
ECL585 (*arcA1*)	S	S

a. R and S denote resistance and sensitivity respectively to toluidine blue and methylene blue

4.10 Discussion

In this chapter the mapping and characterisation of *scrP* has been described. It codes for SCRP-27A, one of four proteins purified by Fujita *et al* (1987) and Ueshima *et al* (1992) which cross react with polyclonal antibodies originally raised against sub-region 2.2 of *E. coli* sigma 70 and sigma 32. From its DNA sequence, *scrP* does not appear to encode a protein related to the sigma 70 or sigma 54 families of transcription initiation proteins, and SCRP-27A does not bind to RNA polymerase core *in vitro* (Ueshima *et al*, 1992). SCRP-27A appears to be inessential for *E. coli* cell viability under a variety of growth conditions.

The *scrP* gene has here been unambiguously mapped between *EcoRV* and *Pvu*II sites near 3416 kb, just to the left of *arcB*. This has been verified by DNA sequencing. *rpoN* and 3 associated open reading frames have been physically mapped by others on the lambda 522/523 overlap near 4315kb, to the left of *scrP* and *arcB* (Merrick *et al* 1993). The corresponding position of *scrP* on the *E. coli* K12 genetic map is approximately 69.75 minutes. This agrees with the cotransduction, data which placed *scrP* at a distance of almost 0.7 minutes from *argG* which itself maps at 69 minutes on the chromosome. However when compared with the most recent version of the genetic map (Bachman 1990) anomalies are apparent in the order of the *rpoN*, *scrP-arcB* and *gltBDF* genes. As figure 4.17 shows, *rpoN* is placed to the right of *gltBDF* on the genetic map, near 70 minutes. This map position is

tentative, in that the position of *rpoN* with respect to other known markers is unclear. However the unequivocal location of *rpoN* on the physical map near the leftward limit of the lambda 522/523 overlap (Merrick *et al* 1993) clearly suggests that its actual location is to the left of *scrP*, *arcB* and *gltBDF*. In confirmation, *gltBDF* has been clearly located within the overlap between the Kohara phages lambda 523 and lambda 524 near 3425 kb on the physical map and thus to the right of *scrP* and *arcB* (summarised in figure 4.17B). Similarly, the previous genetic mapping of *arcB* to the right (clockwise) of *gltBDF* (Iuchi and Lin 1989) should now be reconsidered in the light of the accurate physical location of the respective genes.

Figure 4.17

Locations of the *arcB* and *scrP* genes on the *E. coli* genetic map. The map is based on the 1987 version of the *E. coli* genetic map (Bachman 1987). The scales show co-ordinates on the genetic map in minutes. Relevant genetic markers are shown and their orientations indicated by arrows where they are known.

A. The order of the *gltBDF*, *arcB* and *rpoN* loci (after Iuchi *et al* 1989). Figures in brackets, indicated by vertical arrows below the map, show the map positions (in minutes) of *gltBDF* and *rpoN* on the 1990 version of the genetic map (Bachman 1990).

B. The location of *scrP* on the genetic map, and the revised order of the *gltBDF*, *arcB* and *rpoN* loci based on their respective locations on the physical map of the *E. coli* K12 chromosome (Kohara *et al* 1987).

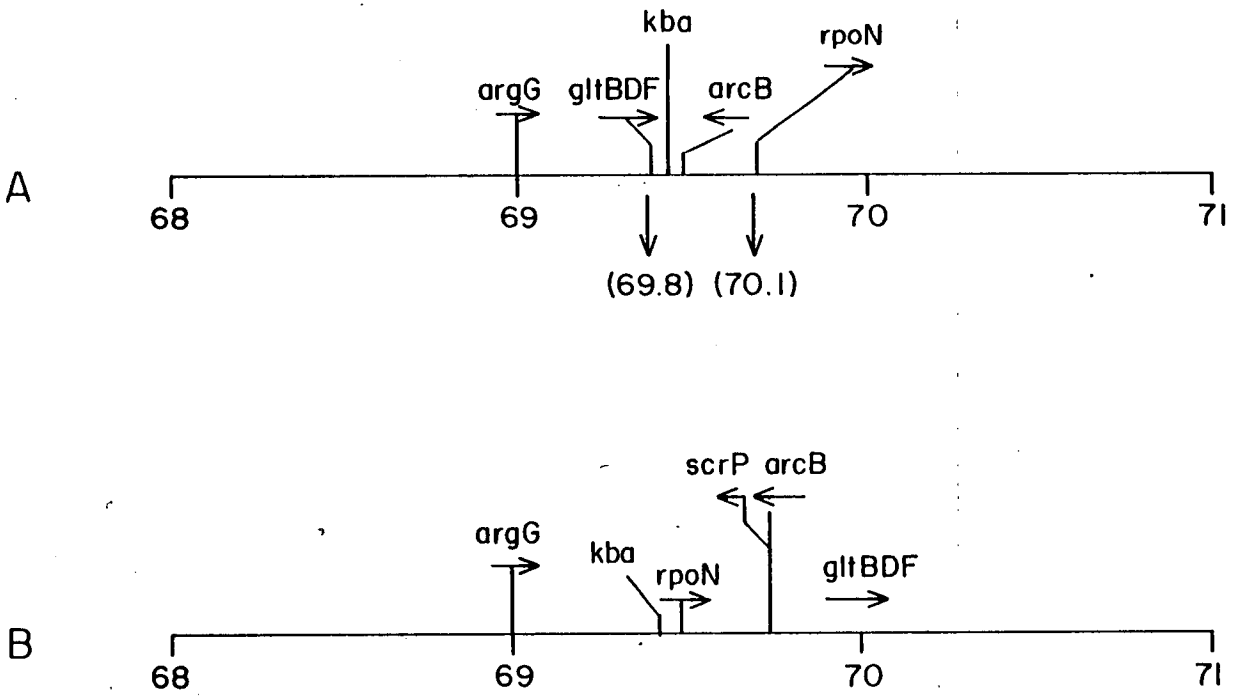

A
kba
rpoN
argG
gltBDF
arcB
68
69
70
71
(69.8)
(70.1)
B
scrP arcB
argG
kba
rpoN
gltBDF
68
69
70
71

FIGURE 4.17

The direction of transcription of *scrP* was determined counter-clockwise, that is from right to left on the physical map, by hybridisation studies using single stranded mp18-15 and mp19-23 DNAs and the oligonucleotide probe scrp27A-1. This conclusion was verified by the *scrP* nucleotide sequence, which is encoded 250 bp downstream of and in the same direction as *arcB*. The latter was already thought to be transcribed in the counter-clockwise direction, based on comparisons of the restriction sites apparent within the *arcB* sequence and on the physical map of the *E. coli* K12 chromosome (Iuchi and Lin 1990). The transcriptional orientation of *arcB* and *scrP* is away from *oriC*, in the direction of DNA replication, a characteristic often associated with strongly expressed *E. coli* genes (Brewer 1988, Medigue *et al* 1993). However the relatively low incidence of optimal codons (61%) within the *scrP* coding sequence is more characteristic of moderately to weakly expressed *E. coli* genes; comparable, for example, with *trpA* (Fop = 0.61) encoding a component of tryptophan synthetase, and *thrS* (Fop = 0.71) whose product is present at a level of ~1000 molecules per *E. coli* genome (Pedersen *et al* 1970). In contrast heavily expressed *E. coli* genes such as *lpp*, encoding murein lipoprotein, and *rplL* encoding the ribosomal protein L7/L12 have very high Fop values, of 0.98 and 0.96 respectively (Ikemura 1986). The *lpp* and *rplL* gene products are present within the cell at ~1.5×10^5 and ~4.5×10^4 molecules per genome, respectively (Pedersen *et al* 1978). Coincidentally *arcB* is another example of an *E. coli* gene that is not strongly

expressed (Medigue *et al* 1993) but is transcribed in the same direction as the DNA is replicated.

Considering the codon usage data discussed above, and the negative results obtained in the immunoprecipitation experiment described in Section 4.5 it seems unlikely that the ~28 kDa polypeptide generated by pFMT1 in minicells is SCRP-27A. The polypeptide is efficiently expressed and may be encoded by uncharacterised DNA situated downstream of *scrP* in pFMT1.

This might be investigated further by analysis of polypeptides encoded by pDSA28 which carries the Tn*901 kan* gene inserted into the *scrP* coding sequence. If as suspected the ~28 kDa protein is not *scrP* then it should be generated intact by pDSA28. Further minicell expression studies using smaller sub-clones covering the *arcB-scrP* locus and downstream DNA might prove useful in locating the genes encoding the 28 kDa and 14 kDa polypeptides produced by pFMT1.

The apparent failure of pFMT1 to express SCRP-27A lends support to the tentative conclusion, derived from the transcription analyses described in Section 4.7, that *scrP* may be co-transcribed with *arcB*. The *arcB* promoter and a proximal portion of the *arcB* gene are missing from pFMT1. Accordingly *scrP* would not be transcribed in this plasmid, providing that *arcB* and *scrP* are obligatorily co-transcribed. A plasmid carrying the whole of *arcB* and *scrP* might generate SCRP-27A which could be identified by expression studies in minicells, combined with immunoprecipitation of plasmid encoded proteins

using anti-2.2 peptide antibodies. Preparation of materials for such experiments is underway in this laboratory.

Returning to the primer extension data presented in Section 4.7.1, the original purpose of this experiment was to investigate whether the putative promoter identified shortly upstream of *scrP* is actually responsible for its transcription. From the data obtained this does not appear to be so, at least under the conditions employed. It remains possible that this potential promoter is functionally significant under different growth conditions.

The majority of *scrP* transcription appears to be initiated far upstream of the gene, probably from the *arcB* promoter recently identified by Iuchi and Lin (1992). At this stage, however, the co-transcription of *arcB* and *scrP* may only be tentatively suggested. Confirmation could be obtained for example by transcriptional fusion studies, and by detailed S1 nuclease analysis of mRNA synthesised *in vivo*, to further investigate whether any transcription termination and/or initiation occurs within the intergenic region. In relation to this, Iuchi and Lin (1990) identified a sequence immediately downstream of the *arcB* gene which they speculated might be a transcription terminator. The sequence is shown below:

ACTAAAAAATGACCCCGGCTAGACCGGGGT GCGCGAATA

Whilst it is probable that the underlined sequences will form a G:C-rich stem-loop RNA structure, which might cause RNA polymerase to pause, the region entirely lacks the U-rich

sequence downstream which would be required to convert the pause into factor-independent termination of transcription (d'Aubenton-Carafa *et al* 1990). It is true that some terminators also have an A-rich sequence immediately upstream of the stem-loop (Wright *et al* 1992). However when that is present in isolation (as here) it is unlikely to cause termination in the forward direction, although termination of transcription converging on *arcB* could be predicted. More generally, factor-dependent transcriptional terminators cannot readily be recognised by sequence examination. S1 nuclease analysis could determine whether termination occurs at any such sites, provide confirmatory data regarding the apparent inactivity of the putative *scrP* promoter, and examine the possibility of transcription in the opposing direction.

A strong feature of figure 4.11 is the large number of terminations occurring throughout the length of the gel, indicating that reverse transcription could proceed all the way into the downstream end of *arcB*. Reverse transcriptase is known to pause and/or terminate transcription at sites of strong secondary structure, or of course where RNA has been cleaved by ribonucleases. The large number of reverse transcript termini may suggest that the mRNA between *arcB* and *scrP* is highly structured. Equally it is possible that at least the more prominent termination sites upstream of *scrP* might represent processing of the mRNA by endoribonucleases. In *E. coli* RNase III, RNase E and RNase M are all capable of cleaving and processing mRNA molecules and often this plays a role in the

regulation of gene expression. For example cleavage of a site between the ribosomal *rplKAJL* operon and the *rpoBC* genes, which are co-transcribed, is important for the translational efficiency of the *rpoBC* genes (Malloch 1990), whilst cleavage of *lacZYA* mRNA affects transcript stability and hence the level of each protein expressed (Cannistraro *et al* 1986).

A major stumbling block in attempts to elucidate a possible function for *scrP* is the lack of sequence similarity between SCRP-27A and proteins of known function and the absence of any sequence motifs of known function and/or structure. It is extremely unlikely that SCRP-27A is a member of the sigma 70 or sigma 54 family. It shows no significant similarity with any of the 4 conserved regions common to the sigma 70 family. In this respect it resembles other SCRPs identified by Fujita *et al* (1987). Therefore *scrP* was disrupted by insertion of a kanamycin resistance gene into its coding sequence. The *scrP :: kan* mutation does not appear to affect the ability of *E. coli* to grow under a variety of conditions between 30°C and 46°C. Additionally the *scrP :: kan* strains DSA100 and DSA101 can be stored on LB or Spizizen minimal agar plates at 4°C for several weeks without significant loss of viability when compared to their $ScrP^+$ parents. Clearly *scrP* is not essential for *E. coli* viability under standard growth conditions or in stationary phase. One possibility which must be considered is that the *scrP :: kan* disruption does not completely destroy SCRP-27A function. That is, it is possible that a partially or fully active truncated and/or chimaeric protein may be

produced by the mutant allele. Whilst it is not possible to exclude this, it is unlikely. In support of this, during the mapping of *arcB* Iuchi and Lin (1989) generated a *rpoN-arcB* deletion which should also have deleted *scrP*, as it is located between those genes. Their deletion strain was completely viable and exhibited only the phenotypes associated with *rpoN* and *arcB* mutations, namely requirement for exogenous nitrogen, and derepression of the Arc modulon during anaerobiosis.

As discussed above, there is some evidence to suggest that *scrP* may be cotranscribed with *arcB*. Obviously this raises the possibility that SCRP-27A may participate in the functioning of the Arc modulon. The dye resistance of DSA100 and DSA101 does not exclude this and further studies will be required in order to fully investigate any role played by SCRP-27A in the ArcAB regulatory system. In particular the effects of the *scrp :: kan* mutation on the levels of enzymes under positive **and** negative regulation by the ArcAB system should be studied. In this respect an *sdh-lacZ* operon fusion has been obtained in order to investigate the effects of *scrP :: kan* on negative regulation of *sdh* during anaerobiosis. Similarly *cyd-lacZ* fusions might be used to assay for effects of *scrP :: kan* on positive control by ArcAB.

Obviously SCRP-27A may not be involved in any way in the regulation of the *arc* modulon. Therefore the latest edition of the *E. coli* linkage map (Bachman 1990) was examined for

possible candidates for the *scrP* locus which might have been mapped under a different name. Only 3 loci which have not been physically mapped or sequenced are shown near 69 minutes and to the left of *gltBDF*. These are *kba* (located near 69.5 minutes), *azaB* (69.7 minutes) and *dgd* (68. ± 1 minute). Of these, *kba* encoding ketos-bisphosphate-aldolase which is required for the metabolism of galacticol (dulcitol) (Lengler 1977) has been eliminated. DSA100 and DSA101 were both able to grow on galactitol as the sole carbon source at 30°C. As expected neither the *scrP :: kan* strains nor their parent strains were able to grow on galactitol at 42°C as the *kba* gene product is temperature sensitive.

Wu (1976) described a locus named *dgd*, encoding a D-galactose dehydrogenase, which is constitutively synthesised in an *E. coli* mutant that can utilise D-arabitol as sole carbon source. The enzyme was not synthesised in wild type *E. coli* K12 strains, which cannot normally grow on D-arabitol. I believe it is unlikely that *scrP* and *dgd* are the same gene. SCRP-27A is synthesised in exponentially growing *E. coli* K12 and is easily purified and separated on SDS-Polyacrylamide gels. In contrast the *dgd* gene product is synthesised only in *dgd* strains, requires NAD for stability and cannot be purified by SDS PAGE. In the latter case the protein was observed to precipitate within the wells at the top of the gel (Wu 1976).

At present, *azaB* cannot be ruled out as a candidate gene

for SCRP-27A. Mutations in *azaB* confer resistance to azaserine at a high level in strains which carry additional mutations in *aroP* and *azaA* (Williams *et al* 1980). Phenylalanine transport is also affected by mutations in *azaB* alone. Williams *et al* (1980) mapped *azaB* to the right of *gltB* by P1 transduction, but Bachman (1990) has placed it to the left of that marker on the linkage map despite the absence of additional literature describing *azaB* since Williams (1980).

Obviously comparing uptake of radiolabelled phenylalanine into *scrP :: kan* and *scrP*$^{+}$ strains would verify whether *scrP* and *azaB* are the same gene. Alternatively, the *scrP :: kan* mutation should confer intermediate level resistance to azaserine upon its introduction into an *aroP* strain.

No other candidate loci for SCRP-27A were identified on the genetic map. *scrP* may represent a previously unidentified genetic locus on the *E. coli K12* chromosome. Assignment of a function to SCRP-27A will require further characterisation of the protein and perhaps identification of related proteins in other organisms.

CHAPTER 5

FINAL DISCUSSION

It is now clear that neither of the two SCRPs investigated in this study are likely to be new *E. coli* sigmas. However all four SCRPs originally identified by Fujita *et al* (1987) specifically cross reacted with the anti-2.2-peptide antiserum.

Ueshima *et al* (1992) have since discovered that the antiserum contains at least three different species of antibodies which separately recognise sigma 70 and the SCRPs.

By performing immunological competition experiments, using a set of overlapping hexameric peptides corresponding to sub-fragments of the original tetradecamer, they found that both SCRP-23 (C22 protein) and SCRP27-A, as well as the other two known SCRPs (thioredoxin reductase and ribosomal protein S2), were outcompeted by a single hexameric peptide of the sequence GLMKAV. A less well characterised SCRP of ~30 kDa was outcompeted by a second hexamer, EGNIGL, whilst none of these peptides could outcompete sigma 70. Ueshima *et al* (1992) concluded that all four major SCRPs carried epitopes which resemble the peptide GLMKAV, whilst sigma 70, which was outcompeted, like the SCRPs, by the tetradecameric peptide, was presumed to carry a conformational epitope consisting of discontinuous aa residues within region 2.2, or alternatively a linear epitope(s) corresponding to boundary regions between the hexameric sub-peptides. From these data it is clear that sigma 70 is recognised by a different species of antibody than are the major SCRPs.

Comparative sequence analyses (data not shown) failed to identify any regions in the C22 protein or SCRP27-A which resemble significantly region 2.2 of sigma 70 and in particular the sequence GLMKAV. Ueshima *et al* (1992) similarly found no significant similarities between region 2.2 and three of the SCRPs (thioredoxin reductase, ribosomal protein S2 and the C22 protein). The exact nature of the epitopes carried by the SCRPs remains unclear, although it seems unlikely that simple linear epitopes are responsible for their cross-reactivity with the antipeptide antiserum.

Obviously this approach has been unsuccessful in identifying new *E. coli* sigmas. However whether *E. coli* contains yet more sigmas is still a valid question. It is now clear that sigmas play a major role in bacterial gene regulation and a large number of sigmas from diverse bacteria have been characterised at the DNA and aa sequence level. The identification of conserved regions, common to many sigmas and their division into sub-families (Lonetto *et al* 1992) has now made it possible to screen DNA libraries using probes specific for regions which are highly conserved amongst members of a sub-family. Indeed such an approach, using oligonucleotide probes, is underway in this laboratory and will hopefully prove more successful than that described in this work.

APPENDIX

PUBLISHED WORK

JOURNAL OF BACTERIOLOGY, June 1992, p. 3826–3827
0021-9193/92/113826-02$02.00/0

Vol. 174, No. 11

Locations of Genes Encoding Alkyl Hydroperoxide Reductase on the Physical Map of the *Escherichia coli* K-12 Genome

DAVID A. SMILLIE,[1] RICHARD S. HAYWARD,[1]* TAEKO SUZUKI,[2] NOBUYUKI FUJITA,[2] AND AKIRA ISHIHAMA[2]

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland, United Kingdom,[1] *and National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan*[2]

In previous work an antiserum raised against a synthetic tetradecapeptide whose sequence is common to the major and heat shock sigma subunits of RNA polymerase was used to identify sigma cross-reacting proteins (SCRPs) in *Escherichia coli* (6). Four of these were purified for further analyses, which established that they were unlikely to be alternative sigma subunits (15). Indeed, amino-terminal sequencing of the four SCRPs, followed by data base comparisons (15), led to the almost certain identification of three of these proteins as thioredoxin reductase, ribosomal protein S2, and the small subunit of alkyl hydroperoxide reductase. The genes encoding the first two of these, *trxB* and *rpsB*, have already been mapped on the chromosome of *E. coli* K-12 (1). The third protein (SCRP-23) has an amino-terminal sequence (15) which differs in only 1 of 30 positions from that of the C22 component of *Salmonella typhimurium* alkyl hydroperoxide reductase (14; in the *Salmonella* protein the 23rd residue, an isoleucine, is conservatively replaced by a valine). In both organisms alkyl hydroperoxide reductase consists of two components: the 22-kDa polypeptide C22 (encoded by *ahpC*) and the 52-kDa *ahpF* product, F52a (3, 4, 7, 13). It is thought to function in the detoxification of lipid and other hydroperoxides that are produced during oxidative stress (for example, within the phagolysosomes of macrophages). In both bacterial species alkyl hydroperoxide reductase is induced in response to low levels of hydrogen peroxide, through the positive regulatory action of the *oxyR* gene product (3, 4, 13). The *S. typhimurium ahpC* and *-F* genes have been mapped (13) and sequenced (14). They lie between *ent* and *lip* at approximately 13 min on the genetic map and are transcribed in the same direction (with *ahpC* upstream). Although separated by 249 bp, they are thought to be cotranscribed.

In the *E. coli* chromosome *ahpC* and *-F* are linked to *entA* (map position 13.85 min [1]), as in *S. typhimurium* (13). However, they have not been mapped physically, nor are their positions marked on the latest genetic map of *E. coli* K-12 (1). In order to map *ahpC* accurately, we prepared a mixed-sequence heptadecanucleotide (a mixture of 64 species) which would encode amino acids 13 through 18 (QAFKNG) of SCRP-23. We used this mixture to probe membranes carrying an ordered DNA library, derived from lambda vectors carrying physically mapped fragments of the *E. coli* K-12 chromosome (8, 11).

The miniset phages λ166 and λ167 were consistently found to give strong positive signals. They were amplified by growth on the *recD* strain NM621 (16), and their DNAs were subjected to Southern hybridization analysis (12) using the mixed-sequence probe. The positively hybridizing region was unambiguously identified as lying between the *BglI* sites located close to 653 and 654 kb on the physical map (Table 1). Thus, *ahpC* is located close to 14.2 min on the genetic map (1), between *entA* and *rna* (Table 1).

This location of *ahpC*, which is wholly compatible with the results of the earlier work of Storz et al. (13), has been further confirmed by DNA sequencing (in progress). The sequencing work has also revealed that as in *S. typhimurium*, the homolog of the *ahpF* gene lies downstream of *ahpC* and both genes are transcribed in the same (clockwise) direction (14).

We have subcloned a 4.1-kb *Hin*dIII-*Kpn*I fragment of *E. coli* DNA encompassing the above-described region from λ166 (8), together with a *Kpn*I-*Bgl*II linker fragment, into pAC184 (2), between the *Hin*dIII and *Bam*HI targets of the vector. The resulting plasmid, pDSA23, was introduced by transformation into the minicell-producing strain DS410 (5). When minicells derived from the resulting Cm^r transformants were labelled with [^{35}S]methionine and their proteins were fractionated on sodium dodecyl sulfate-polyacrylamide gels (12), strong synthesis of a polypeptide with an estimated size of 22.5 kDa was observed, together with moderately strong synthesis of a 53-kDa protein (data not shown). We believe that these are the *ahpC* and *ahpF* products, respectively. Production of polypeptides of the same sizes, and in similar proportions, was also observed in UV-irradiated *E. coli* S159 or S159(λ*ind*) cells (10), provided that they were

TABLE 1. Physical and genetic features and coordinates in the *aphCF* region of the *E. coli* K-12 chromosome

Physical map coordinate (kb)[a]	Gene or other feature	Genetic map position (min)[b]
623	*entD*	13.3
	entA	13.85
644	Left end of λ166 insert	
652.7	Left end of λ167 insert	
652.9	*Bgl*I target	
653.2[c]	Start of *ahpC* gene	14.2 [c]
653.75[c]	End of *ahpC* gene	
653.95	*Bgl*I target	
654[c]	Start of *ahpF* gene	14.25[c]
659[d]	*rna*	14.4
662	Right end of λ166 insert	
670	Right end of λ167 insert	

[a] Unless otherwise indicated, these are derived as accurately as possible from reference 8.
[b] Unless otherwise stated, these are taken from reference 1.
[c] Determined in the present work.
[d] Taken from reference 9.

* Corresponding author.

infected with either λ166 or λ167. As expected, the 22.5-kDa polypeptide encoded by pDSA23 in minicells was selectively precipitable by the antibody used in the original isolation of SCRP-23 (6). Control minicells from a DS410 strain containing pAC184 displayed no detectable synthesis of the 22.5- or 53-kDa polypeptide.

We conclude that *ahpC* and *ahpF* are orientated clockwise, with the latter gene downstream, between positions 652 and 656 on the physical map (8), and therefore near min 14.2 (between *entA* and *rna*) on the genetic map (1), of the chromosome of *E. coli* K-12.

This work was funded by the Wellcome Trust (United Kingdom), the Ministry of Education, Science and Culture of Japan, and (for a working visit to Mishima by R.S.H.) the British Council and the Royal Society (United Kingdom).

We thank S. Crosthwaite and I. Oliver for advice on the minicell procedure, M. Gallagher for discussion, and J. Ferguson for other help.

REFERENCES

1. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
2. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. **134**: 1141–1156.
3. **Christman, M. F., R. W. Morgan, F. S. Jacobson, and B. N. Ames.** 1985. Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in *Salmonella typhimurium*. Cell **41**:753–762.
4. **Demple, B., and J. Halbrook.** 1983. Inducible repair of oxidative DNA damage in *Escherichia coli*. Nature (London) **304**:466–468.
5. **Dougan, G., and D. Sherratt.** 1977. The transposon Tn*1* as a probe for studying ColE1 structure and function. Mol. Gen. Genet. **151**:151–160.
6. **Fujita, N., A. Ishihama, Y. Nagasawa, and S. Ueda.** 1987. RNA polymerase sigma-related proteins in *Escherichia coli*: detection by antibodies against a synthetic peptide. Mol. Gen. Genet. **210**:5–9.
7. **Jacobson, F. S., R. W. Morgan, M. F. Christman, and B. N. Ames.** 1989. An alkyl hydroperoxide reductase from *Salmonella typhimurium* involved in the defense of DNA against oxidative damage. Purification and properties. J. Biol. Chem. **264**:1488–1496.
8. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
9. **Meador, J., III, and D. Kennell.** 1990. Cloning and sequencing the gene encoding *Escherichia coli* ribonuclease I: exact physical mapping using the genome library. Gene **95**:1–7.
10. **Newman, A. J., T. G. Linn, and R. S. Hayward.** 1979. Evidence for co-transcription of the RNA polymerase genes *rpoBC* with a ribosomal protein gene of *Escherichia coli*. Mol. Gen. Genet. **169**:195–204.
11. **Noda, A., J. B. Courtright, P. F. Denor, G. Webb, Y. Kohara, and A. Ishihama.** 1991. Rapid identification of specific genes in *E. coli* by hybridization to membranes containing the ordered set of phage clones. BioTechniques **10**:474–477.
12. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Storz, G., F. S. Jacobson, L. A. Tartaglia, R. W. Morgan, L. A. Silveira, and B. N. Ames.** 1989. An alkyl hydroperoxide reductase induced by oxidative stress in *Salmonella typhimurium* and *Escherichia coli*: genetic characterization and cloning of *ahp*. J. Bacteriol. **171**:2049–2055.
14. **Tartaglia, L. A., G. Storz, M. H. Brodsky, A. Lai, and B. N. Ames.** 1990. Alkyl hydroperoxide reductase from *Salmonella typhimurium*. Sequence and homology to thioredoxin reductase and other flavoprotein disulfide oxidoreductases. J. Biol. Chem. **265**:10535–10540.
15. **Ueshima, R., N. Fujita, and A. Ishihama.** Identification of *Escherichia coli* proteins crossreacting with antibodies against region 2.2 peptide of RNA polymerase sigma subunit. Biochem. Biophys. Res. Commun., in press.
16. **Whittaker, P. A., A. J. B. Campbell, E. M. Southern, and N. E. Murray.** 1988. Enhanced recovery and restriction mapping of DNA fragments in a new lambda vector. Nucleic Acids Res. **16**:6725–6736.

BIBLIOGRAPHY

d'Aubenton Carafa, Brody, E., & Thermes, C. (1990) J. Mol. Biol. **216**: 835-858

Amabile-Cuevas, C.F., & Demple, B. (1991) Nucl. Acids Res. **19**: 4479-4484

Andrews, S.C., Harrison, P.M. and Guest J.R. (1991) J. Gen. Microbiol. **137**: 361-367

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

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

Bachman, B. (1987) In Neidhardt, F.C., & Van Bogelen, R.A. (ed.) *Escherichia coli* and *Salmonella typhimurium*: Cellular and Mollecular Biology, American Society for Microbiology, Washington, D.C., pp 807-876

Bachman, B.J. (1990) Microbiol. Rev. **59**: 130-197

Ballongue, J., Amine, J., Masion, E., Petitdemange, H., & Gay, R., (1986), Biochimie **68**: 575-580

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

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

Bowen, S.W., & Hassan, H.M. (1988) J. Biol. Chem. **260**: 992-925

Brewer, B. (1988) Cell **53**: 679-686

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

Buckae B. (1993) Mol. Microbiol. **9**: 671-680

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

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

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

Buxton, R.S., Drury, L.S., & Curtis, C.A.M., (1983), J. Gen. Microbiol. **129**: 3363-3370

Cannistraro, V.J., & Kennel, D. (1986) J. Mol. Biol. **192**: 257-274

Carlioz, A. & Touati, D. (1986) EMBO J. **5**: 623-630

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

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

Chang, B.Y., & Doy, R.H. (1993) Biochem. 294: 43-47

Chen, Y.F., & Helmann, J.D. (1992) Proc. Natl. Acad. Sci. USA 89: 5123-5127

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

Christman, M.F., Morgan, R.W., Jacobson, F.S., & Ames, B.N. (1985) Cell 41: 753-762

Christman, M.F., Storz, G., & Ames, B.N. (1989) Proc. Natl. Acad. Sci. USA 86: 3484-3488

Chui, J.T., Loewen, P.C., Switala, J., Gennis, R.B., & Timkovich, R. (1989) J. A . Chem. Soc 111 7046-7050

Claiborne, A., & Fridovich, I. (1978) J. Biol. Chem. 254: 4245-4252

Claiborne, A., Ross, R.P., & Parsonage, D. (1992) TIBS 17: 183-191
83-186

Collins, J.F., Coulson, A.F.W., & Lyall, A. (1988) CABIOS 4: 67-71

Craig, E.A., & Gross, C.A. (1991) TIBS 16: 135-140

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

Davies, K.J., & Lin, S.W. (1988) Free. Radic. Biol. Med. 5: 215-223

Demple, B. (1991) Annu. Rev. Genet. 25: 315-317

Demple, B. & Halbrook, J. (1983) Nature (London) 304: 446:448

Demple, B., Johnson, A., & Fung, D. (1986) Proc. Natl. Acad. Sci. USA 83: 7731-7735

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

Dougan, G., & Sherratt, D. (1977) Mol. Gen. Genet. 151: 151-160

Eisenstark, A. (1989) Adv. Genet. 26: 99-147

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

Erickson, J.W., Vaughn, V., Walter, W.A., Neidhart, F.C. & Gross,

C.A. (1987) Genes and Dev. 1: 419-432

Espinosa-Urgel, M.E., & Tormo, A. (1993) Nucl. Acids Res. 21: 36667-3670
Farr S.B., Natvig, D.O., & Kogoma, T. (1985) J. Bacteriol. 164: 1309-1316

Farr, S.B., D'Ari. R., & Touati, D. (1986) Proc. Natl. Acad. Sci. USA 83: 8268-8272

Farr, S.B., Kogoma, T. (1991) Microbiol. Rev. 55: 561-585

Farr, S.B., Touati, D. & Kogoma, T. (1988) J. Bacteriol. 170: 1837-1842

Feinberg, A.P., & Vogelstein, P. (1984) Anal. Biochem. 137: 266-267

Flores, B.M., Batzer, M.A., Stein, M.A., Petersen, C., Diedrich, D.L., & Torian, B.E., (1993) Mol. Microbiol 7: 755-763

Francis, K.P., & Gallacher, M.P. (1993) Infect. Immun. 61: 640-649

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

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

Gardner, P.R., & Fridovich, I. (1991) J. Biol. Chem. 266: 1478-1483

Gardner, P.R., & Fridovich, I. (1991) J. Biol. Chem. 266: 19328-19333

Georgopolous, C., & Welch, W.J. (1993) Annu. Rev. Cell. Biol. 9: 610-634

Georgopolous, C., Ang, D., Liberek, K., & Zylicz, M.,Morimoto, R.I., (1990), In Stress Proteins in Biology and Medicine, Morimoto, R., Tissieres A., & Georgopolous C., (ed.) Cold Spring Harbor Press, Cold Spring Harbor, N.Y., pp 191-222

Gesteland, R.F.(1966) J. Mol. Biol. 16: 67-84

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

Gillen, K.L., & Hughs, K.T., (1991), J. Bacteriol. 173: 2301-2310

Gillin, F.D., & Diamond, L.S. (1981,a). Exp. Parasitol. 52: 9-17

Gillin, F.D., & Diamond, L.S., (1981,b). Exp. Parasitol. 51: 382-391

Gillin, F.D., Reiner, D.S., Levy, R.B., & Henhart, P.A. (1984) Mol. Biochem. Parasitol. **13**: 1-12

Glass, R.E., & Hayward, R.S. (1993) World Journal of Microbiology & Biotechnology **9**: 403-413

Goldmark, P.J., Curry, J., Morns, C.F., & Walker-Simpson, M.K. (1992) Molecular Biology **19**: 433-441

Greenberg J.T., & Demple, B. (1986) J. Bacteriol. **168**: 1026-1029

Greenberg J.T., & Demple, B. (1988) EMBO J. **7**: 2611-2617

Greenberg J.T., & Demple, B. (1989) J. Bacteriol. **171**: 3933-3939

Greenberg J.T., Chou, J.H., Monach, P., & Demple, B. (1991) J. Bacteriol. **173**: 4433-4439

Greenberg J.T., Monach, P., Chou, J.H., Josephy, P.D., & Demple, B. (1990) Proc. Natl. Acad. Sci. USA **87**: 6181-6185

Greer, S., & Perham, R.N. (1986) Biochem. **25**: 2736-2742

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

Gross, C.A., Strauss, D.B. and Erickson, J.W. (1990) In R.I. Morimoto, A. Tissieres and C. Georgopolous (ed.) Stress Proteins in Biology and Medicine, pp 167-189. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

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

Hassan, H.M., & Sun, H.C.H. (1992) Proc. Natl. Acad. Sci. USA **89**: 3217-3221

Hassan, H.M., & Fridovich, I. (1977) J. Biol. Chem. **252**: 7667-7672

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

Hayward, R.S. and Scaife, J.G. (1976) Nature **260**: 646-648

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

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

Helmann, J.D., & Chamberlin, M.J. (1987) Proc. Natl. Acad. Sci. USA **84**: 6422-6424

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

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

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

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

Henikoff, S., Haughn, G.W., Calvo, J.M., & Wallace, J.C. (1988). Proc. Natl. Acad. Sci. USA **85**: 6602-6606

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

Hoerter, J., Eisenstark, A., & Touati, D. (1989) Mutat. Res. **215**: 161-165

Igarashi, K., Fujita, N., & Ishihama A. (1990). Nucl. Acids Res. **18**: 5945-5948.

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

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

Imlay, J.A., & Fridovich, I. (1991) Free Radical Research Communications **12**: 59-66

Imlay, J.A., & Fridovich, I. (1992) J. Bacteriol. **174**: 953-961

Imlay, J.A., & Lin, S. (1986) Science **240**: 1302-1309

Ishihama, A. (1992) Mol. Microbiol. **6**: 3283-3288

Ito, K., Egawak., & Nakamura, Y. (1991) J. Bacteriol. **173**: 1492-1501

Iuchi, S., & Lin, E.C.C. (1988) Proc. Natl. Acad. Sci. USA **85**: 1888-1892

Iuchi, S., Cameron, D., & Lin, E.C.C. (1989) J. Bacteriol. **171**: 868-873

Iuchi, S., Matsuda, Z., T Fujiwara, & Lin, E.C.C. (1990) Mol. Microbiol. **4**: 715-727

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

Jacobson, F.S., Morgan, R.W., Christman, M.F., & Ames, B.N. (1989) J. Biol. Chem. **264**: 1488-1496

Jaskunas, S.R., Lindhal, L., & Nomura, M. (1975), Proc. Natl. Acad. Sci. USA **72**: 6.

Jones, C.H., & Moran Jr, C.P., (1992) Proc. Natl. Acad. Sci. USA **89**: 1958-1962

Jordan, S.R., & Pabo, C.O. (1988) Science **242**: 893-899.

Kappus, H. (1985) In H. Sies (ed) Oxidative Stress, Academic Press Inc., New York, pp 273-310

Kappus., & Seis, H. (1981) Experientia **37**: 1233-1258

Kashlev, M., Lee, J., Zalenskaya, K., & Goldfarb, A. (1990) Science **248**: 1006-1009

Kenney, T.J., York, K., Youngman, P., & Moran Jr, C.P. (1989) Proc. Natl. Acad. Sci. USA **86**: 9109-9113

Kindler, P., Keil, T.U., & Hofschneider, P.H. (1973) Mol. Gen. Genet. **126**: 53-59.

Kohara, Y. (1990) In K. Drlica & M. Riley (ed.), The Bacterial Chromosome, American Society for Microbiology, Washington, D.C.

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

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

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

Kyte, J., and Doolittle, R.F. (1982) J. Mol. Biol. **157**: 105-132

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

Lange, R., & Hengge-Aronis, R. (1991) Mol. Microbiol. **5**: 49-59

Lengler, J. (1977) Mol. Gen. Genet. **152**: 83-91

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

Lipinska, B., Fayet, O., Baird, L., & Georgopoulos, C. (1989) J. Bacteriol. **171**: 1574-1584

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

Loewen, P.C. (1984) J. Bacteriol. **157:** 622-626

Loewen, P.C., & Switala, J. (1986) J. Biol. Chem. **64:** 638-646

Loewen, P.C., Triggs, B.L., George, C.S., & Hrabarchuk, B.E. (1985) J. Bacteriol. **162:** 661-667

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

Loprasert, S., Negoro, S., & Okada, H. (1989) J. Bacteriol. **171:** 4871-4875

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

Malloch, R.A. (1990) Ph.D. Thesis, University of Edinburgh

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

Mathieu, I., Meyer, J., & Moulis, J.M. (1992) Biochem. J. **285:** 255-262

Matin, A. (1991) Mol. Microbiol. **5:** 3-10

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

Meador III J., & Kennel, D. (1990) Gene **95:** 1-7

Medigue, C., Viari, A., Henaut, A., & Danchin, A. (1993) Microbiol. Rev. **57:** 623-654

Meek, D.W., & Hayward, R.S. (1986) Mol. Gen. Genet. **202:** 500-508

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

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

Merrick, M., Jones, D.H.A., & Thomas, C.M. (1993) J. Bacteriol. **175:** 1548-1549

Merrick, M., & Chambers, S. (1992), J. Bacteriol. **174:** 7221-7226

Messing, J., & Vierra, J., (1982) Gene **19:** 259-268

Miller. J. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.

Moran, C.P. Jnr. (1989) in I. Smith, R. Selpecky and P. Setlow (ed.) Regulation of Prokaryotic Development, pp 167-184. American Society for Microbiology, Washington, D.C.

Morgan, R.W., Christman, M.F, Jacobson, F.S., Storz, G., & Ames, B.N. (1986) Proc. Natl. Acad. Sci. USA **83**: 8059-8063

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

Nagai, H., Yano, R., Erickson, J.W., & Yura, T. (1990) J. Bacteriol. **172**: 2710-2715

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

Nahlik, M.S., Fleming, T.P., & McIntosh, M. (1987) J. Bacteriol. **169**: 4163-4170

Neidhardt, F.C., & Van Bogelen, R.A. (1987) In Neidhardt, F.C. (ed.) *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, American Society for Microbiology, Washington, D.C., Volume 2, pp 1334-1345

Nemoto, Y., Yamamoto, T., Takada, S., Matsui, Y., & Obinata, M. (1990) Gene **91**: 261-265

Niimura, Y., Ohnishi, K., Yarita, Y., Hidaka, M., Masaki, H., Uchimura, T., Suzuki, H., Kozaki, M. and Uozumi, T. (1993) J. Bacteriol. **175**: 7945-7950

Nikaido, H., & Vaara, M. (1987) In F.C. Neidhardt (ed.) *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, American Society for Microbiology, Washington, D.C. pp 7-22

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

O'Toole, P.W., Logan, S.M., Kostrzynsko, M., Wadstrom, T., Trust, T.J. (1991) J. Bacteriol. **173**: 505-513

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

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

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

Ovchinnikov, Y.A., Monastyrskaya, G.S., Gubanov, V.V., Guryev, S.O.,Salomatina, I.S., Shuvaeva, T.M., Lipskin, V.M., & Sverdlov, E.D. (1982) Nucl. Acids Res. **10**: 4035-4044

Pedersen, S., Bloch, P.L., Reeh, S., & Neidhardt, F.C. (1978) Cell **14**: 179-190

Post, L.E., & Nomura, M. (1979) J. Biol. Chem. **254**: 10604-10606.

Reed, S.L., Flores, B.M., Batzer, M.A., Stein, M.A., Stroeher, V.L., Carlton, J.E., Diedrich, D.L., & Torian, B.E. (1992) Infect. Immun. **60**: 542-549

Rowland, G.C., & Glass, R.E. (1990) Bioessays **12**: 343-346.

Rudd, K.E. (1992) In J. Miller (ed.) A Short Course in Bacterial Genetics: A Laboratory Manual & Handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Russell, M., & Model, P. (1988) J. Biol. Chem. **263**: 9015-9019

Sak, B.D., Eisenstark, A., & Touti, D. (1989) Proc. Natl. Acad. Sci. USA **86**: 3271-3275

Sak, B.D., Gralla, J.D. (1990) Cell **62**: 945-954

Salser, W., Gesteland, R.F., & Bolle, A. (1967) Nature **215**: 588-591

Sambrook, J. Fritsch, E.F., & Maniatus, T. (1989) Molecular Cloning A Laboratory Manual 2nd Edition. Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklens, S., & Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**: 5463-5467

Sasse-Dwight, S., & Gralla, J.D. (1988) Proc. Natl. Acad. Sci. USA **85**: 8934-8938

Sasse-Dwight, S., & Gralla, J.D. (1990) Cell **62**: 945-954

Saucier, J., & Wang, J. (1972) Nature New Biol. **239**: 167-170.

Seeman, N.C., Rosenberg, J.M., & Rich, A. (1976) Proc. Natl. Acad. Sci. USA **73**: 804-808

Shapira, S.K., Chou, J., Richaud, F.V., & Casadaban, M.J. (1983) Gene **25**: 71-82.

Siegele, D.A., Hu, J.C., Walter, W.A., & Gross, C.A. (1989) J. Mol. Biol. **206**: 591-603

Simpson, R.B. (1987) Cell **18**: 277-285.

Spiro, S., & Guest, J.R. (1991) TIBS 16: 310-314

Stadtman, T.C. (1965) Non-Heme Iron Proteins: Role in Energy Conservation. (San Pietro, E., ed.) pp 439-445, Antioch Press, Yellow Springs, Ohio.

Stoker, N.G., Pratt, J.M., & Holland, I.B. (1984) Transcription & Translation, a Practical Approach, Hames B.D. & Higgins, S.J. (ed.) 153-176

Storz, G. Christman, M., Sies, H., & Ames, B.N. (1987) Proc. Natl. Acad. Sci. USA **84**: 8917-8921

Storz, G. Jacobson, F.S., Tartaglia, L.A., Morgan, R.W., Silveria, L.A. & Ames, B.N. (1989) J. Bacteriol. **171**: 2049-2055

Storz, G. Tartaglia, L.A., & Ames, B.N. (1990) Science **248**: 189-194

Storz, G. Tartaglia, L.A., Farr, S.B., & Ames, B.N. (1990) TIG **6**: 363-368

Stragier, P., Parsot, C., & Bouvier, J. (1985) FEBS Lett. **187**: 11-15.

Straus, D.B., Walter, W.A., & Gross, C.A. (1987) Nature **329**: 343-350

Straus, D.B., Walter, W.A., & Gross, C.A. (1989) Genes & Dev. **3**: 2003-2010

Su, W., Porter, S., Kustu S., & Echols, H. (1990) Proc. Natl. Acad. Sci. USA **87**: 5504-5508

Summerfield, F.W., & Tappel, A.L. (1983) Biochim. Biophys. Acta. **740**: 185-189

Takao, M., Oikawa, A., & Yasui, A. (1990) Arch. Biochem. Biophys. **185**: 473-482

Tanaka, K., Shiina, T., & Takahashi, H. (1988) Science **242**: 1040-1042.
Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A., & Takahashi, H. (1993) Proc. Natl. Acad. Sci. USA **190**: 3511-3515.

Tao, K., Makino, K., Yonei, S., Nakata, A., & Shinagawa, H. (1991) J. Biochem. **109**: 262-266

Tao, K., Makino, K., Yonei, S., Nakata, A., Shinagawa, H. (1989) Mol. Gen. Genet. **218**: 371-376.

Tao, K., Fujita, N., & Ishihama, A. (1993) Mol. Microbiol. **7**: 859-864.

Tardat, B., & Touati, D. (1993) Mol. Microbiol. 9: 53-63.

Tartaglia, L. A., Storz, G., & Ames, B.N., (1989) J. Mol. Biol. 210: 709-719.

Tartaglia, L.A., Gimeno, J., Storz, G., & Ames, B.N. (1992) J. Biol. Chem. 267: 2038-2045.

Tartaglia, L.A., Storz, G., Brodsky, M.H., Lai, A., & Ames, B.N., (1990) J. Biol. Chem. 265: 10535-10540.

Torian, B.E., Flores, B.M., Stroeher, V.L., Hagen, F.S., & Stamm, W.E., (1990), Proc. Natl. Acad. Sci. USA 87: 6358-6362.

Touati, D., (1988), J. Bacteriol. 170: 2511-2520.

Triggs-Raine, B.L., Doble, B.W., Mulvey, M.R., Sorby, P.A. & Loewen, P.C., (1988), J. Bacteriol. 170: 4415-4419.

Tsaneva, I.R., & Weiss, B. (1990) J. Bacteriol. 172: 4197-4205.

Ueshima, R., Fujita, N., & Ishihama, A. (1992) Biochem. Biophys. Res. Commun. 184: 634-639.

Van Bogelen, R.A., Kelley, P.M., & Neidhardt, F.C. (1987) J. Bacteriol. 169: 26-32.

Von Ossowksi, I., Mulvey, M.R., Leco, P.A., Bonys, A., & Loewen, P.C. (1991) J. Bacteriol. 173: 514-520.

Voordouw, G. (1988) Gene 69: 75-83.

Waldburger, C., Gardella, T., Wong, R., & Susskind, M.M. (1990) J. Mol. Biol. 215: 267-276.

Walkup, L.K.B., & Kogoma, T. (1989) J. Bacteriol. 171: 1476-1484

Wang, Q., & Kaguni, J.M. (1989) J. Bacteriol. 171: 4248-4253.

Wedel, A., Weiss, D.S., Popham, D., Droge, P., & Kustu, S. (1990) Science 248: 413-520.

Weiner, L., Brissette, J.L., & Model, P. (1991) Genes & Dev. 5: 1912-1923.

Wellman, A., & Meares, C.F. (1991) Biochem. Biophys. Res. Commun. 177: 140-144.

Whitaker, P.A., Campbell, A.J.B., Southern, E.M., & Murray, N.E. (1988) Nucl. Acids. Res. 16: 6725-6736.

Williams, M.V., Kerr, T.J., Lemmon, R.D., & Tritz, G.J. (1980) J. Bacteriol. 143: 383-388.

Wright, J.J., (1987), Ph.D. Thesis, University of Edinburgh

Wright, J.J., Kumar, A., & Hayward, R.S. (1992) EMBO J. 11 1957-1964

Wu, J., & Weiss, B. (1992) J. Bacteriol. **174**: 3915-3920.

Wu, T.T. (1966) Genetics **54**: 405-410.

Wu, T.T. (1976) J. Gen. Microbiol. **94**: 246-256.

Xu, X., Koyama, N., Cui, M., Yamagishi, A., Nosho, Y., Oshima, T. (1991) J. Biochem. **109**: 678-683.

Xu, X., Kanaya, S., Koyama, N., Sekiguchi, T., Nosho, Y., Ohashi, S.,& Tsuda, K. (1989) J. Biochem. **105**: 626-632.

Yallaly, P., & Eisenstark, A. (1990) Biochem. Biophys. Res. Commun. **169**: 64-69.

Yamaguchi, R., Matsuo, K., Yamakazi, A., Takahashi, M., Fukasawa, Y., Wada, M., & Abe, C. (1992) Infect. Immun. **60**: 1210-1216.

Yamamoto, T., Matsui, Y., Natori, S., & Obinata, M. (1989) Gene **80**: 337-343.

Yanisch-Perron, C., Viera, J., Messing, J. (1985) Gene **33**: 103-119

Yura, T. & Ishihama, A. (1979) Annu. Rev. Genet. **13**: 59-97.

Zillig, W., Zechel, K., Rabussay, D., Schachner, M., Sethi, Y.S., Palm, P., Heil, A. & Seifert, W. (1970) Cold Spring Harbor Symp. on Quant. Biol. **35**: 47-58.

Zuber, P., Healy, J., Carter III, H.L., Cutting, S., Moran Jr. C.P., & Losick, R. (1989) J. Mol. Biol. **206**: 605-614.