Molecular organization and functional analysis of the CFA/II CS3 region of Enterotoxigenic *Escherichia coli*

Meachery Bhaskaran Jalajakumari,

B. Sc. (Ag.), M. Sc. (Ag.) (Kerala, INDIA)

Thesis submitted for the degree of Doctor of Philosophy

in

The University of Adelaide

(Faculty of Science)

Department of Microbiology and Immunology **July, 1992**

Dedicated to the loving memory of my father

"Yadä - yadä hi dharmasya glänir bhavati bhäratha
Abhyutthänam adharmasya tadätmanam srjämy aham
Paritränaya sadhunäm vinasaya ca duskrtäm
Dharma samsthäpanarthäya sambhavämi yuge - yuge."

- The Bhagavad Gita

ABSTRACT

Enterotoxigenic *Escherichia coli* (ETEC) have two major virulence determinants:- the pilus/fimbria required for colonization and the enterotoxin which induces the diarrhoea. Various types of pili are found in human ETEC, and the coli surface antigen 3 (CS3) is common to strains bearing Colonization Factor Antigen II (CFA/II). This factor is poorly characterised at the level of genetic organisation and regulation of its synthesis. This thesis describes the molecular organisation and regulation, and suggests possible roles, of the various genes and their products in the biosynthesis of CS3. The studies reported here have all been performed in *E. coli* K-12, using a pBR322-based clone containing the plasmid encoded genes from the wild type CFA/II ETEC strain, PB176.

Minicell and *in vitro* analyses of translation and DNA sequencing showed that all genes involved in the biosynthesis of CS3 are present on a minimal fragment of 4.6 kb, as previously described by Manning *et al.* (1985). Although the DNA sequence analysis confirmed some observations made in the above paper, it also required the number and sizes of polypeptides to be revised. Specifically, the previously reported size of a 97 kDa polypeptide has been revised to 82 kDa and three smaller polypeptides of sizes 31, 22 and 9 (kDa), respectively, have been identified. Eight ORFs with predicted sizes of 82, 61, 46, 31, 27, 22, 17 and 9 kDa, respectively, are found in a uniquely overlapping arrangement, in that, four of them are fully contained within, and share the same reading frame with, another. These five ORFs each initiated with a different AUG codon but terminated with the rarely used UAG termination codon. Experimental evidence for independent initiation of translation to produce these polypeptides, as contrasted to post-translational processing, was obtained by the use of site-directed mutagenesis. Furthermore, oligonucleotide-directed mutagenesis of the common termination codon resulted in lengthening of the polypeptides and loss of pili biosynthesis. However, this effect on pili synthesis was reversed when a frame-shift was introduced near the

termination codon. Translation to produce the 22 kDa polypeptide required the suppression of an inframe UAG codon and it was found that insertion of glutamine at this position, by the *supE* mutation in *E. coli* K-12, was critical for CS3 biogenesis.

The open reading frame (ORF) of the major pilin subunit (CS3) was identified by determining the N-terminal amino acid sequence of purified pilin. An alternative signal peptide cleavage site for the CS3 polypeptide, used preferentially in *E. coli* K-12, has been identified.

Amino acid sequence comparisons revealed significant homologies between the 82 kDa and several putative channel proteins involved in other known ETEC pili systems. Similarly, the 27 kDa protein showed homologies to other periplasmic transport proteins. No significant homology between the CS3 polypeptide and other pilin subunits was found but the 22 kDa polypeptide showed some homology to CS3, suggesting that it participated directly in pili formation.

Cell fractionation studies showed that the 27 kDa polypeptide was located in the periplasmic fraction. A large amount of the CS3 pili subunit was also seen in the periplasmic fraction. All other proteins, except the 22 kDa protein which was observed in very small amounts in minicell translations, were in the membrane fraction. An accumulation of the 22 kDa protein was observed when pilus biosynthesis was abolished by introducing an inframe deletion in a region shared by the 82 kDa and 61 kDa polypeptides.

The mRNA analyses revealed at least four transcription units and four promoters have been shown to be functional *in vivo*. The CS3 ORF is preceded by its own promoter but is also encoded on the other three transcripts. Neither the transcription nor translation of the CS3 polypeptide required functional products of any of the other proteins. Expression of pili on the cell surface, however, required all proteins.

An IS*91* element is found in the DNA, upstream to the cloned CS3 biosynthesis genes. The promoter region of the 27 kDa protein showed homology to the promoter

region of the *cfaA* gene described by Hamers *et al.* (1989). Similarity to a sequence downstream to the coding region of the adhesin F17-G gene was also found near the P_{27} promoter. A putative IHF binding site precedes the translation initiation site of *cp82*. No Rho-independent terminator was found on the sequence, but comparison of transcription patterns in *E. coli* K-12 and PB176 indicated that termination of transcription occurred in the area following the *cpCS3* coding region.

A model for CS3 biosynthesis has been proposed in the thesis. This model assumes the 82 kDa protein as a channel protein, the 27 kDa protein as a periplasmic transporter and the 22 kDa protein as a pili-associated basal protein. The remaining five proteins are membrane-associated but are of unknown functions.

Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or any other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying.

M. B. Jalajakumari. July, 1992.

Acknowledgements

I want to thank the following people for their help and support during the course of this work.

Professors D. Rowley and C. Burrell, and Dr. I. Kotlarski for permitting me to work in the Department of Microbiology and Immunology.

Professor P. A. Manning for his expert supervision and constant encouragements. I want to thank him specially for his very thorough and critical reading of this thesis manuscript. I also want to express my gratitude for his understanding nature and the many personal helps. Thank you, Paul.

I thank Dr. M. Heuzenroeder for his supervision during the early part of the project and all friends in Lab. 432, past and present, for their suggestions, discussions and lots of fun.

I am grateful to Chris Cursaro and Garry Penney for technical assistance, Terry Proctor for his excellent photography, and Miss Angelo Parker, Dr. Connor Thomas and Mrs. Elena Voss for their help in electron microscopy.

Many thanks to the friends and colleagues, not mentioned by names, in the department of Microbiology and Immunology, for providing a pleasant and stimulating atmosphere to work in.

I thank Andy Cheel in the Information Technology Divison for kindly permitting the use of computing equipments for printing this thesis.

Miss Betty Brockelbank, Mr. Chris Clark, Drs. Renato Morona, Sue Willams and Steve Attridge are gratefully ackowledged for critically reading the manuscript. I want to thank Betty especially for her love and support during the entire period of my research.

Finally, I want to thank my husband, Siv, for carrying out expert computer analyses and extensive corrections of my writing. I also want to thank him for his patience during the preparation of this thesis, and for helping to maintain my sanity.

LIST OF ABBREVATIONS

aa	aminoacid
Ap	Ampicillin
Ap^R	Ampicilin resistant
APS	Ammonium persulphate
ATP	Adenosine 5'-triphosphate
BCIG	5-bromo-4-chloro-3-indolyl-3-β-d-Galactopyranoside
bp	base or nucleotide pair
BSA	bovine serum albumin
CFA	colonization factor agar.
CFB	colonization factor broth.
Cm	Chloramphenicol.
cpm	counts per minute
dNTP	deoxyribonucleoside triphosphate
ddNTP	dideoxyribonucleoside triphosphate
DTT	dithiothreitol
EtBr	ethidium bromide
g	gram
Gal	galactose
g/l	grams/litre
IPTG	Iso-propyl β-D-thiogalactopyranoside
kb	kilobases
Kcal	kilocalories
kDa	kilodalton
Klenow	Klenow fragment of *E. coli* DNA polymerase I
Km	kanamycin

mA	milliAmps
mg	milligram
ml	millilitre
µg	microgram
µl	microlitre
mm	millimeter
mM	millimolar
ng	nanogram
nmol	nanomol
nt	nucleotide
OD_{600}	optical density at 600 nm
RF	replicative form
r.p.m	revolutions per minute
TEMED	N,N,N',N'- tetramethylenediamine
Tn	transposon
Tris	Tris-hydroxymethyl-aminomethane
X-gal	5-bromo-4-chloro-3-indoyl-β-d-galacto-pyranoside

CONTENTS

Abstract i

Statement iv

Acknowledgements v

Abbreviations vi

List of tables xv

List of figures xvi

CHAPTER 1: General Introduction

1.1 Introduction 1

1.2 Enterotoxigenic *E. coli* (ETEC) 2
- 1.2.1 Enterotoxins 2
- 1.2.1.1 LT toxin 3
- 1.2.1.2 ST toxins 5
- 1.2.2 Fimbrial Adhesins 6

1.3 E. coli fimbriae 8
- 1.3.1 Type I (MSHA) fimbriae 9
- 1.3.2 MRHA fimbriae 9
- 1.3.2.1 MRHA fimbriae of uropathogenic *E. coli* 9
- 1.3.2.2 MRHA fimbriae of Enterotoxigenic *Escherichia coli* 12

1.4 Morphology and structure of fimbriae. 14

1.5 Structural aspects of the fimbriae. 16

1.6 Biogenesis of fimbriae 19
- 1.6.1 Pap fimbrial system 19
- 1.6.2 Type I fimbriae 23
- 1.6.3 F4 (K88) fimbriae 25
- 1.6.4 K99 fimbriae 27

1.7 Fimbrial biogenesis in Human ETEC 29
- 1.7.1 General aspects 29
- 1.7.2 Colonization Factor Antigen I (CFA/I) 29
- 1.7.3 Colonization Factor Antigen II (CFA/II) 32
- 1.7.3.1 CS1 and CS2 33
- 1.7.3.2 Rns protein 36
- 1.7.3.3 CS3 37
- 1.7.4 Colonization Factor Antigen III (CFA/III) 39
- 1.7.4 Colonization Factor Antigen IV (CFA/IV) 39

	1.7.6	PCFO9	42
	1.7.7	Other human ETEC	42
1.8	**Aims of this thesis**		43

CHAPTER 2: Materials And Methods

2.1	**Bacterial strains**		45
2.2	**Bacteriophage strains**		45
2.3	**Plasmids**		45
2.4	**Media and Buffers**		45
	2.4.1	CFA broth	45
	2.4.2	L Broth	45
	2.4.3	Nutrient Broth	46
	2.4.4	2 x YT Broth	46
	2.4.5	M13 minimal medium	46
	2.4.6	TFB I (Transformation Broth I)	46
	2.4.7	TFB II	47
	2.4.7.1	Transformation component #a	47
	2.4.7.2	Transformation component #b	47
	2.4.8	CFA agar plates	47
	2.4.9	M13 minimal plates	47
	2.4.10	Nutrient Agar plates	47
	2.4.11	McConkey agar plates	48
	2.4.12	YT Soft agar	48
2.5	**Buffers**		48
	2.5.1	10 x TM	48
	2.5.2	10 x TE	48
	2.5.3	10 x TAE	49
	2.5.4	10 x TBE	49
	2.5.5	TBS for Western blotting	49
	2.5.6	TTBS for Western blotting	49
	2.5.7	Transfer buffer for Western blotting	49
	2.5.8	10 X PBS	50
	2.5.9	TMACl solution	50
2.6	**Chemicals**		50
2.7	**Enzymes**		51
2.8	**Radionucleotides**		51
2.9	**Miscellaneous materials**		52
2.10	**Synthesis of oligodeoxynucleotides**		52

2.11	**General techniques**		52
	2.11.1	Maintenance of bacterial strains	52
	2.11.2	Lyophilization of bacterial cultures	52
	2.11.3	Purification of colonies and plaques	52
2.12	**Analysis and manipulation of DNA**		53
	2.12.1	Small scale plasmid DNA extraction	53
	2.12.1.a	Method 1	53
	2.12.1.b	Method 2	53
	2.12.1.c	Method 3	54
2.13	**Large scale plasmid DNA isolation**		54
	2.13.1.a	Method 1	54
	2.13.1.b	Method 2	55
2.14	**Phenol extraction of nucleic acids**		55
	2.14.a	Equilibration of phenol	55
	2.14.b	Extraction with phenol	56
2.15	**Ethanol precipitation**		56
2.16	**DNA quantitation**		57
2.17	**Restriction Endonuclease Digestion**		57
2.18	**Size determination of DNA fragments**		58
2.19	**Analytical and preparative separation of DNA fragments**		58
	2.19.a	Method 1	58
	2.19.b	Method 2	58
	2.19.c	Method 3	59
2.20	**Polyacrylamide gel electrophoresis**		59
	2.20.1.a	Non-denaturing PAGE	59
	2.20.1.b	Denaturing (sequencing) gel	60
2.21	**Elution of DNA from polyacrylamide gel**		60
2.22	**Dialysis** 60		
2.23	**De-phosphorylation of DNA**		61
	2.23.a	Method 1	61
	2.23.b	Method 2	61
2.24	**End-labelling and end-filling**		61
2.25	**Ligation**		62
2.26	**Transformation of E. coli**		62
	2.26.a	$CaCl_2$ Method	62
	2.26.b	$RbCl_2$ Method	63

2.27	**M13 transfection**		63
2.28	**Mobilization of plasmid into ETEC cells**		64
2.29	**Labelling of the oligonucleotides**		64
2.30	**Galactokinase Assay**		65
2.31	**Analysis and manipulation of RNA**		65
	2.31.1	Preparation of cellular RNA	65
	2.31.1.a	Method 1	65
	2.31.1.b	Method 2	66
	2.31.2	Quantitation of RNA	67
	2.31.3	Preparation of single stranded probe	67
	2.31.4	Hexamer labelling method to make double strand probe	67
	2.31.4	Preparation of RNA Dot Blots	67
	2.31.5	Northern blot	68
	2.31.5.a	Glyoxalation of RNA	68
	2.31.5.b	Electrophoresis	68
	2.31.5.c	Transfer to the nitrocellulose filter	68
	2.31.5.d	Pre-hybridization	68
	2.31.5.e	Hybridization, washing of filters and autoradiography	69
	2.31.6	Primer extension	69
2.32	**Protein analysis**		69
	2.32.1	Minicell labelling	70
	2.32.2	Cell fractionation	70
	2.32.3	Prokaryotic In vitro transcription/translation	71
	2.32.4	SDS-Polyacrylamide Gel Electrophoresis of proteins	71
	2.32.5	Autoradiography	72
	2.32.6	Preparation of CS3 antisera	72
	2.32.7	Crude pili preparation	72
	2.32.8	Colony transfer and blotting with antiserum	72
	2.32.9	Western Blotting	73
	2.32.9.a	Using peroxidase	74
	2.32.9.b	Using Di-aminobenzidine (DAB)	74
	2.32.10	Electronmicroscopy	74
	2.32.10.a	Preparation of protein A-Gold	74
	2.32.10.b	Immuno-gold labelling of whole cells	74
2.33	**M13 cloning and sequencing**		75
	2.33.1	Generation of Subclones for sequencing	75
	2.33.1.a	DNaseI deletion method	75
	2.33.1.b	Restriction fragment subcloning method	76
	2.33.2	Preparation of M13 replicative form (RF) DNA	76
	2.33.3	Purification plaques for sequencing	76
	2.33.4	Screening M13 clones for inserts	77
	2.33.5	Single stranded phage preparation	77

2.33.6	Single stranded DNA preparation	78
2.33.7	Sequencing	78
2.33.7.1	Dideoxy sequencing protocol with Klenow fragment	78
2.33.7.1.a	Annealing of primer and template	78
2.33.7.1.b	Polymerization reaction	78
2.33.7.1.c	Fixing of gel and autoradiography	79
2.33.7.2	Dideoxy sequencing protocol with Sequenase	79
2.33.7.3	Double stranded DNA sequencing (CsCl DNA)	81
2.33.7.4	Double stranded DNA sequencing (miniprep DNA)	81
2.34	Site-directed oligonucleotide mutagenesis	82
2.34.1	Method 1	82
2.34.2	Method 2	82

CHAPTER 3: Sequence Analysis Of The DNA Encoding The CS3 Gene Cluster

3.1	**Introduction**	84
3.2	**Results**	85
3.2.1	pPM484 encodes all genes required for CS3 pilus synthesis in *E. coli* K-12	85
3.2.2	The nucleotide sequence of pPM484	86
3.2.3	Analysis of the nucleotide sequence	86
3.2.4	Definition and summary of ORFs	87
3.2.4.1	*cp27*	88
3.2.4.2	*cp82*	89
3.2.4.3	*cp22*	90
3.2.4.4	*cpCS3*	90
3.2.5	Overlapping translation is needed for other ORFs	91
3.2.5.1	*cp61*	92
3.2.5.2	*cp46*	92
3.2.5.3	*cp31* and *cp9*	92
	cp31	92
	cp9	93
3.2.6	Other Features of the sequence	93
3.2.6.1	Sequence homologies with other DNAs	93
3.2.6.1a	IS*91* element	94
3.2.6.1b	*cfaA*	94
3.2.6.1c	*E. coli* adhesin gene, F17-G	94
3.2.6.2	Promoters	95
3.2.6.3	Terminators	95
3.2.6.4	Direct repeats	95
3.2.6.5	Protein binding sites	95
3.2.6.6	G+C content and Restriction Enzyme Sites	96
3.3	**Discussion**	97

CHAPTER 4: Identification, Analyses And Roles Of The CS3 Pilus Biosynthesis Proteins

4.1	**Introduction**		101
4.2	**Results**		102
	4.2.1	Correlation of protein bands to ORFs.	102
	4.2.1.1	Cp27	102
	4.2.1.2	Cp82	104
	4.2.1.3	Cp61	105
	4.2.1.4	Cp46	106
	4.2.1.5	Cp31	106
	4.2.1.6	Cp22	107
	4.2.1.7	Cp9	108
	4.2.1.8	CS3	108
	4.2.2	Termination codon of the overlapping ORFs	109
	4.2.3	Involvement of the Cp proteins in pilus biosynthesis.	110
	4.2.3.1	Cp27	110
	4.2.3.2	Cp82	110
	4.2.3.3	Cp61 and Cp46	111
	4.2.3.4	Cp31 and Cp9	111
	4.2.3.5	Cp22	112
	4.2.3.6	CS3	112
	4.2.4	Structural properties of the proteins	112
	4.2.4.1	Cp27	112
	4.2.4.2	Cp82	113
	4.2.4.3	Cp61, Cp46, Cp31, Cp22 and Cp9	113
	4.2.4.4	CS3	113
	4.2.5	Leader peptides and membrane spanning regions of the Cp proteins.	114
	4.2.6	Potential roles of proteins.	115
	4.2.7	Protein similarities with other pilus biosynthesis genes.	115
	4.2.8	Intracellular localisation of the gene products	116
4.3	**Discussion**		117
	4.3.1	Analysis of translational initiation	117
	4.3.2	Mutational analyses to identify potential roles	118
	4.3.3	Identification of potential roles by cell fractionation	119

CHAPTER 5: Transcriptional Analysis Of The CS3 Operon

5.1	**Introduction**		122
5.2	**Results**		122
	5.2.1	Growth conditions for studying the transcription patterns	122
	5.2.2	Northern blot analysis of transcripts	123

	5.2.3	Localisation of P_{27} and P_{CS3} by primer extension	125
	5.2.4	*In vivo* analyses of promoter activities	125
	5.2.5	Estimation of relative promoter strengths by GalK assay	126
5.3	**Discussion**		127

CHAPTER 6: Translational Control Of CS3 Pilus Biosynthesis

6.1	**Introduction**		129
6.2	**Results**		129
	6.2.1	Suppression of UAG_{3578} is required for CS3 biosynthesis	129
	6.2.2	Mutating UAG_{3578} to a sense codon restored pilus biosynthesis	130
	6.2.3	Mutating UAG_{3523} to a sense codon resulted in loss of pilus biosynthesis	131
	6.2.4	Suppression of UAG_{3578} by supF did not restore pilus biosynthesis	131
6.3	**Discussion**		132

CHAPTER 7: General Discussion

7.1	**Introduction**	134
7.2	**Gene organization within the CS3 operon**	135
7.3	**Possible significance of overlapping genes**	136
7.4	**Pattern of Transcription**	139
7.5	**Codon Usage**	140
7.6	**Read-through Translation**	141
7.7	**Properties of proteins**	143
7.8	**Cellular location of gene products**	144
7.9	**Model for CS3 pilus biosynthesis**	146
7.10	**Conclusion**	148

REFERENCES 151

List of Tables

Table 1.1	Properties of ETEC fimbriae
Table 1.2	Summary of human fimbrial polypeptides
Table 1.3	N-terminal amino acid sequences of fimbrial subunits
Table 1.4	N-terminal amino acid sequences of fimbrial subunits in human ETEC
Table 2.1	Bacterial Strains
Table 2.2	Vectors and Plasmid Clones
Table 2.3	Sequencing Reagents
Table 3.1	Summary of ORFs
Table 3.2	Summary of potential promoters
Table 3.3	Summary of Relevant Restriction sites
Table 4.1	Summary of protein compositions
Table 5.1	Galactokinase Assay to detect promoter activities
Table 7.1	Codon frequencies in ETEC, E. coli K-12 and the cp genes

List of Figures

Figure 3.1 Western blot of crude pili from pPM474 vs pPM484

Figure 3.2 Electron micrographs showing cell-surface CS3 pili

Figure 3.3 Minicell translation of pPM484 and pBR322

Figure 3.4 a. Nucleotide sequence of the HindIII fragment, encoding the CS3 pilus biosynthesis genes

b. Sequencing strategy

Figure 3.5 Codon usage plot of the DNA sequence

Figure 3.6 Schematic diagram of ORFs and promoters.

Figure 3.7 DNA sequence homologies

Figure 3.8 Transcriptional terminator in the CS3 operon.

Figure 4 .1 Comparison of the in vitro and minicell translation of pPM484

Figure 4 .2 Minicell translation in the presence of 8% ethanol

Figure 4 .3A Minicell translation of the mutant derivatives of pPM484

Figure 4.3B Minicell translation of the *EcoRI* and the *Bgl*II deletion mutants

Figure 4 .4 Effect of ApaI deletion on Cp22 and CS3.

Figure 4 .5 Minicell labelling of the UAG_{3523} mutant

Figure 4 .6 Restriction analysis to verify the UAG_{3523} mutation.

Figure 4 .7 Western blot of pili from cells carrying the mutant plasmids

Figure 4 .8 Immunogold EM of cells carrying mutant derivatives of pPM484

Figure 4 .9 Secondary structure and hydrophobicity profiles of Cp27

Figure 4 .10 Secondary structure and hydrophobicity profiles of Cp82

Figure 4 .11 Secondary structure and hydrophobicity profiles of Cp61

Figure 4 .12 Secondary structure and hydrophobicity profiles of Cp46

Figure 4 .13 Secondary structure and hydrophobicity profiles of Cp31

Figure 4 .14 Secondary structure and hydrophobicity profiles of Cp9

Figure 4 .15 Secondary structure and hydrophobicity profiles of Cp22

Figure 4 .16 Secondary structure and hydrophobicity profiles of CS3

Figure 4 .17 CLUSTAL alignment of Cp27 and other pili biosynthesis proteins

Figure 4 .18 CLUSTAL alignment of Cp82 and other pili biosynthesis proteins

Figure 4 .19 Homology of Cp22 and CS3 to each other and to other ETEC proteins

Figure 4 .20 Intra-cellular location of proteins

Figure 5 .1 Relative abundance of mRNA in cells grown in different culture media

Figure 5.2 Linear diagram showing probes, putative promoters and transcripts

Figure 5.3 Messenger RNA detected by pPM484-specific probes.

Figure 5.4 Primer extension on P27 transcript

Figure 5.5 Primer extension on PCS3 transcript

Figure 5.6 Detection of promoter activities *in vivo*

Figure 6.1 Nucleotide sequence between UAG_{3523} and UAG_{3578}

Figure 6.2 Immunogold EM of PC2251 (Su^+) and PC2251 (Su^-) minicells

Figure 6.3 Western blot of crude pili from Su^+ and Su^- strains:

Figure 6.4 Western blot of crude pili from *E. coli* K-12 and ETEC strains

Figure 6.5 Immunogold EM of ETEC, mobilized with DNA from pPM484

Figure 6.6 Immunogold EM of UAG mutants

Figure 6.7 Western blot of crude pili from *supF* strain carrying pPM484 DNA

Figure 7.1 Structural organization of genes involved in fimbrial biogenesis

Figure 7.2 Model for CS3 pilus biosynthesis

CHAPTER 1

General Introduction

CHAPTER 1

General Introduction

1.1 Introduction

Escherichia coli is the most common facultative anaerobe found among the flora that inhabit the large intestine, where it plays an important role in maintaining normal physiological functions in humans. However, *E. coli* is also the most common Gram-negative human pathogen, responsible for both intra- and extra- intestinal infections. Within the species there are pathogenic strains that cause distinct syndromes of diarrhoeal diseases.

Pathogenic *E. coli*, involved in distinct disease patterns, belong to a limited number of O-H serotypes, bear distinct capsular antigens (Orskov *et al.*, 1977), and show specific binding properties to epithelial cell surface receptors. Based on virulence properties, interactions with the intestinal mucosa, clinical syndromes, epidemiology and O:H sero-types, the diarrhoeagenic *E. coli* are classified into five major groups (Levine, 1987):

1. Enterotoxigenic *E. coli* (ETEC) - a major cause of travellers' diarrhoea and infant diarrhoea in less developed countries.

2. Enteroinvasive *E. coli* (EIEC) - a cause of dysentery.

3. Enteropathogenic *E. coli* (EPEC) - an important cause of infant diarrhoea.

4. Enterohaemorrhagic *E. coli* (EHEC) - a cause of haemorrhagic colitis and haemolytic uremic syndrome.

5. Enteroadherent *E. coli* (EAEC) - identifiable only by their pattern of adherence to Hep-2 cells in tissue culture.

Since this thesis is an investigation of the biosynthesis of the CS3 pilus belonging to the Colonization Factor Antigen II (CFA/II) of human ETEC, the rest of this chapter concentrates on the toxins and fimbriae in different ETEC strains. Particular attention has

been given to what is known about the components of various fimbrial assembly systems. Emphasis has also been placed upon the biogenesis of cell-surface pili in human and animal ETEC strains.

1.2 Enterotoxigenic *E. coli* (ETEC)

ETEC are considered to be one of the major causes of diarrhoea in infants in developing countries (Black *et al.*, 1981). ETEC also causes diarrhoea in travellers to the third world countries, often referred to as "travellers' diarrhoea", "Matlab Monsoon", "Montezuma's revenge", or "Delhi belly" (Gorbach *et al.*, 1975; Merson *et al.*, 1976). Infection is acquired by ingesting contaminated food or water.

In the veterinary field, ETEC are known to cause severe diarrhoeal disease in young herd animals by producing toxins similar to those of strains infecting humans (Gyles, 1971; Gyles *et al.*, 1974). ETEC are also a significant cause of diarrhoea in neonatal herd animals (Gross *et al.*, 1978; Guinee *et al.*, 1977; Gyles, 1971; Smith and Linggood, 1972). The animal serotypes and fimbrial types are distinct from those causing disease in humans and tend to be species-specific (Soderlind and Mollby, 1979).

The production of virulence factors may be affected by environmental conditions such as iron, osmolarity, temperature and nutritional status, e.g. amino acid availability (Båga *et al.*, 1985; Bortolussi *et al.*, 1983; Calderwood and Mekalanos, 1987; Finley and Falkow, 1989; Jovanovich *et al.*, 1988; Mekalanos, 1985; Roosendaal *et al.*, 1986). The colonization of ETEC occurs by means of various host-specific fimbrial adhesins. Bacterial growth *in situ,* and elaboration of enterotoxin(s), lead to increased water and electrolyte secretion into the intestinal lumen, hence the diarrhoeal response and dehydration. Individual strains of ETEC produce either heat-labile (LT) or heat-stable (ST) enterotoxin or both.

1.2.1 Enterotoxins

The ETEC strains isolated from man and animals produced either or both LT and

ST enterotoxins. The genes encoding these toxins are usually found on transferable plasmids (Sherman *et al.*, 1972; Smith and Linggood, 1971).

1.2.1.1 LT toxin

The heat-labile enterotoxins of *Vibrio cholerae* (Cholera Toxin) and *E. coli* (LT) belong to a family of structurally related proteins that cause diarrhoea in humans and animals. These toxins stimulate production of cAMP in target cells by catalysing the ADP ribosylsation of the $G_{s\alpha}$ subunit of the adenylate cyclase complex in the plasma membrane (Holmes *et al.*, 1990; Middlebrook and Dorland, 1984). The heat-labile enterotoxins are oligomeric proteins composed of a single A polypeptide that is non-covalently bound to a pentameric array of B polypeptides (Gill *et al.*, 1981; Holmes *et al.*, 1990). The B polypeptides mediate binding of the toxins to receptors on the plasma membrane of target cells (Holmes *et al.*, 1990). Fragment A1, derived from the A polypeptide by proteolytic cleavage, and reduction of an intra-chain disulphide bond, is enzymatically active and catalyses the ADP ribosyltransferase reaction.

The *E. coli* - *V. cholerae* enterotoxin family is divided into two serogroups (Finkelstein *et al.*, 1987; Holmes *et al.*, 1986b; Pickett *et al.*, 1986; 1989). Serogroup one consists of cholera toxin (CT), the type I *E. coli* enterotoxins (LT-I; including the variants, LTp-I and LTh-1 produced by *E. coli* strains of porcine and human origin, respectively) and antigenically related enterotoxins from several other Gram-negative enteric bacteria (Calva *et al.*, 1989; Prasad *et al.*, 1990; Rose *et al.*, 1989). Antiserum against CT or LT-I will neutralise the toxicity of other type I toxins (Guth *et al.*, 1986b; Holmes *et al.*, 1986b; Pickett *et al.*, 1986, 1989). Serogroup two consists of the type II *E. coli* enterotoxins (LT-II), including the cross-reacting antigenic variants designated LT-IIa and LT-IIb (Pickett *et al.*, 1986, 1989; Pickett and Holmes, 1990). Antisera to type I enterotoxins do not neutralise the LT-II variants, and *vice versa*.

Comparative genetic and biochemical analyses of these enterotoxins were facilitated by cloning and sequencing the representative operons encoding them and by

purifying the corresponding holotoxins (Dallas and Falkow, 1980; Guth *et al.*, 1986a, 1986b; Holmes *et al.*, 1986a, 1986b, 1988; Pickett *et al.*, 1986, 1987, 1989; Pickett and Holmes 1990). The genes that encode the A polypeptides of type I and II enterotoxins are homologous (Holmes *et al.*, 1990; Pickett *et al.*, 1986, 1987, 1989; Yamamoto *et al.*, 1984a). In contrast, the genes that encode the B polypeptides of type I enterotoxins have little, if any, significant homology with those that encode the B polypeptides of type II enterotoxins (Holmes *et al.*,1988; Pickett *et al.*, 1986, 1989). The B polypeptides of type I enterotoxins differ greatly from those of type II enterotoxins with respect to amino acid sequence, antigenicity and receptor binding specificity (Fukuta *et al.*, 1988; Guth *et al.*, 1986a, 1986b; Holmes *et al.*, 1986b). Nevertheless, type I and type II B polypeptides are approximately equal in size, have a single intra chain disulphide bond between cysteine residues located at nearly identical positions, have similar hydropathy profiles, form pentamers and associate with their corresponding A polypeptides to form holotoxins (Holmes *et al.*, 1990; Yamamoto *et al.*, 1984b).

While both CT and LT-I utilise ganglioside GM-1 as a functional receptor in the plasma membranes, LT-I also binds to intestinal glycoproteins (Griffiths *et al.*, 1986). LT-IIa binds preferentially to ganglioside GD1b, whereas LT-IIb binds most strongly to ganglioside GD1a (Fukuta *et al.*, 1988). It is likely that these gangliosides serve as functional receptors for type II enterotoxins. Both LT-IIa and LT-IIb also bind less strongly to several other gangliosides (Fukuta *et al.*, 1988). The structure of the ganglioside GM1-binding domain of CT and LT-I has been studied by both genetic and biochemical methods (De Wolf *et al.*, 1981; Finkelstein *et al.*, 1987; Ludwig *et al.*, 1985; Tsuji *et al.*, 1985). Connell and Holmes (1992) identified three positions in the mature B polypeptide of LT-IIa at which amino acid substitutions cause loss of ganglioside GDIb-binding activity, and they implicated the hydroxyl groups of threonine at positions 1, 14, and 34 as important for such binding.

1.2.1.2 ST toxins

ETEC produce two kinds of heat-stable enterotoxin (ST) which cause intestinal secretion and diarrhoea (Betley *et al.*, 1986). Two immunologically and genetically distinct forms of ST toxin, ST_A and ST_B, have been cloned and their sequences determined (Lee *et al.*, 1983; So and McCarthy, 1980). The form, ST_A, found in human pathogens causes fluid accumulation in the intestine of infant mice and this is the basis for an effective bioassay (Dean *et al.*, 1972). This form of ST toxin has been termed ST-I to differentiate it from a methanol insoluble ST-II found in some ETEC strains pathogenic for pigs (Lee *et al.*, 1983). The latter, also designated as ST_B by Burgess *et al.* (1978) and ST_P by Olsson and Soderlind (1980), has been primarily associated with ETEC isolated from diarrhoeic swine. Genes encoding this toxin have the highest prevalence among the toxin genes in porcine ETEC (Monckton and Hasse, 1988; Moon *et al.*, 1986). ST-I is well characterised, and the toxic domain involved in the expression has been determined. The heat-stable toxins bear no structural or functional characteristic with LT.

Only ST_A is of relevance to human disease. DNA hybridisation studies, using cloned genes from different sources as probes (Moseley *et al.*, 1982), have shown that at least two distinct sub-classes of ST_A exist. These subclasses, which have been referred to as ST human and ST porcine (or ST bovine), have been shown by DNA sequencing studies to share 69% DNA homology with each other (Moseley *et al.*, 1983; So and McCarthy, 1980). So *et al.* (1976) cloned the gene for ST_A production. Restriction and nucleotide sequence analyses revealed that the ST_A gene resides on a transposon, flanked by two inverted repeats of IS*1* elements (So *et al.*, 1979). This ST_A transposon was termed Tn*1681* and its full nucleotide sequence determined (So and McCarthy, 1980). Depending on the host species of origin, ST_A toxins are either 18 or 19 amino acid, extracellular peptides (Dreyfus *et al.*, 1983), which result from two discrete proteolytic cleavages of a 72 amino acid precursor (Rasheed *et al.*, 1990). Following attachment to a specific receptor located in the intestinal brush border membrane, ST_A

activates the membrane guanylate cyclase and the resulting rise in mucosal cGMP is thought to mediate intestinal ion and fluid transport (Field *et al.*, 1978; Guerrant *et al.*, 1980). Recently, it has been reported that the intestinal guanylate cyclase serves as an ST_A receptor (Schulz *et al.*, 1990). Whether additional ST_A receptors may couple with guanylate cyclase has yet to be determined.

The gene for ST_B has also been cloned and analysed. This toxin originates as a 71 amino acid (Lee *et al.*, 1983; Picken *et al.*, 1983) precursor which is processed in a single step to yield a periplasmic, 48 amino acid species. The latter becomes, without further apparent proteolysis, an extracellular peptide (Kupersztoch *et al.*, 1990). Although ST_B, like ST_A, is a cysteine-rich peptide, the latter is sensitive to trypsin (Whipp, 1987) and does not invoke a cyclic nucleotide response in the gut (Weikel and Guerrant, 1985). It too appears to be associated with a transposable element (Tn*4521*). Dreyfus *et al.* (1991) reported that disulphide bond formation is apparently absolutely required for proper folding and/or export of mature, biologically active ST_B toxin. They also reported that at least two of the amino acids (Arg52 and Asp53) present in a 14 amino acid, closed loop formed by Cys44 and Cys59 are essential for full toxin activity. At present, the mechanism by which ST_B induces intestinal secretion is poorly understood.

1.2.2 Fimbrial Adhesins

ETEC must possess accessory virulence properties in addition to LT and ST toxins to cause diarrhoea. The ability of bacteria to bind to epithelial surfaces is considered to be a pre-requisite for infection. Mechanical protective factors at epithelial surfaces are important in host defences against invading bacteria. These factors include direct removal by washing, e.g. by the flow of urine in the urinary tract, by peristalsis in the intestine and by continuous secretion of saliva in the oral cavity. To colonise epithelial surfaces, bacteria have to overcome the washing effect by firmly attaching to the infection site. This is apparently necessary for the bacteria both to cause

infectious diseases and to establish themselves as members of our normal flora.

The best characterised accessory virulence properties are colonization factors, which allow the bacteria to adhere to specific receptors on enterocytes of the proximal small intestinal mucosa of the host (Levine, 1981). All specific colonization factors so far identified in either animal or human ETEC strains have been shown to be fimbriae (also termed pili).

The term fimbriae, derived from Latin, means "fibres" or "threads", and was introduced as a designation for non-flagellar bacterial appendages by Duguid *et al.* (1955). Fimbriae are supramolecular structures composed of low molecular weight subunits, which are found on the cell surfaces of many pathogenic Gram-negative bacteria (Beachey, 1981). Brinton (1959) later coined the term pili (Latin for hair). Lately, these two terms have generally been used as synonyms. A cell may produce more than one type of fimbria. Some have been shown to mediate attachment to the host cell surface. Bacterial fimbriae are filamentous, non-flagellar surface appendages. They vary in their phylogenetic distribution, morphology, function and number per cell. The terminology and classification of different types of fimbriae is confusing, since there is no general acceptance of a single system of nomenclature. Ottow (1975) proposed a scheme in which six groups were tentatively defined according to the criteria given below.

Group 1 Peritrichous, adhesive organelles found in the *Enterobacteriaceae* (Paranchych and Frost, 1988).

Group 2 Organelles for conjugative genetic exchange (sex pili), which are encoded by plasmid borne genes. These are also the attachment sites for certain bacteriophages (Paranchych and Frost, 1988).

Group 3 Thick (40-60nm wide) hollow tubes found on *Agrobacterium spp.* and other soil bacteria.

Group 4 Fimbriae, defined by their polar location on the cell and their ability to impart a form of bacterial motility, called type 4 or MePhe (because of their

N-terminal amino acid) fimbriae. These appear to function as colonization factors (Elleman and Hoyne, 1984; Paranchych and Frost, 1988).

Group 5 Fimbriae found on cell poles (of certain soil bacteria) but are distinguished by an unusual function, *viz.*, promoting conjugation by effecting bacterial clustering.

Group 6 The bundled filaments of *Corynebacterium renale.*

The terms pilus and fimbria define structures with a specific morphology and do not specify function. Pili that are adhesins are therefore called pilus adhesins. In this thesis, the terms pilus and fimbria will be used as synonyms.

1.3 *E. coli* fimbriae

Adherence of ETEC to the brush borders of epithelial cells of the proximal small intestine is a crucial event in their pathogenesis. A number of fimbrial types can be distinguished within this group. These classifications are based on the specificity of the ligand, the tissue and host infected, location of fimbrial genes (chromosomal or on the plasmid-encoded) and serological variation. (Paranchych and Frost, 1988). A clear correlation has been established between the presence of specific adhesins on the surface of bacteria and their ability to adhere to a given host mucosal surface. These adhesins allow the bacteria to overcome the clearing mechanisms that protect the epithelial surfaces (e.g. peristalsis, mucus secretion), providing a structure which binds to a distinct receptor on the mucosa or in the glycocalyx of epithelial cells (Lark *et al.*, 1986; Smyth, 1986).

Many different species in the *Enterobacteriaceae* possess fimbriae, but most of our knowledge comes from studies of *E. coli*, particularly those strains involved in infectious diseases (Clegg and Gerlach, 1987; Duguid and Old, 1980). They are usually classified functionally by their receptor specificity, as determined by the ability of either fimbriated bacteria or their purified fimbriae to agglutinate erythrocytes from different species. The pili have been divided into two categories, based on haemagglutination

(HA) reactions, mannose sensitive (MSHA) and mannose resistant (MRHA; Ottow, 1975; Pearce and Buchanan, 1980). The sensitivity to mannose derives from the fact that some pili use D-mannose, its derivatives, or D-mannose-containing molecules as receptors (Duguid and Old, 1980; Old, 1972).

1.3.1 Type I (MSHA) fimbriae

MSHA is associated with pili that are often referred to as "type I" or "common pili". These have been observed in a number of different species (Clegg and Gerlach, 1987; Eisenstein, 1988). Type I fimbriae are rigid rods, 7nm in diameter and 1μm in length, distributed uniformly over the cell surface. Their number varies between 100-500 per cell. They mediate characteristic mannose sensitive agglutination of guinea pig erythrocytes. MSHA is displayed by 70%-90% of *E. coli* strains and appears not to be related to bacterial virulence. Type I fimbrial genes in *E. coli* are genetically mapped to distinct loci at 98 minutes on the chromosome (Brinton *et al.*, 1961). Eight of the genes involved in the synthesis and assembly have been cloned and sequenced (Klemm and Christiansen, 1987).

1.3.2 MRHA fimbriae

E. coli strains expressing mannose-resistant fimbriae cause haemagglutination of erythrocytes from certain animal species, thereby providing a simple screening test to detect their presence. MRHA can be mediated by uropathogenic and enterotoxigenic strains of *E. coli*.

1.3.2.1 MRHA fimbriae of uropathogenic *E. coli*.

MRHA is frequently (77%) displayed by acute pyelonephritic isolates, but is less commonly observed with strains isolated from the faecal flora (16%) or cystitis patients (35%; Hagberg *et al.*, 1981). P fimbriae, associated with MRHA, have been designated as *PAP* pili (*P*ili *A*ssociated with *P*yelonephritis), P-fimbriae (Korhonen *et al.*, 1982) or G-fimbriae (Rhen *et al.*, 1986). Apart from their binding specificity, fimbrial adhesins can

be separated into different F groups (F7-F14; Abe, *et al.*, 1987; Orskov and Orskov, 1983, 1985) according to their serologic specificities - e.g. P fimbriae of F13, encoded by Pap operon (Lund *et al.*, 1985; Normark *et al.*, 1983a) and of serotypes $F7_1$ (*fso* determinant; Van Die *et al.*, 1985) and $F7_2$ (*fst* determinant; Van Die *et al.*, 1984a).

The genetic determinants which code for P or P-related fimbriae of serotypes $F7_1$, $F7_2$ and F13 have been mapped respectively to regions next to the *serA* gene (map position 57- 63 minutes), in the vicinity of the *ilv* gene cluster (position 85-90 minutes) and at 95 minutes on the chromosome of O4 and O6 of wild type strains (Hoekstra, *et al.*, 1986; Hull *et al.*, 1986). In contrast, the P-determinant of serotype F8 (*fei*) has been mapped at 17 minutes on the chromosome of an O18:K5 strain (Krallmann-Wetzel *et al.*, 1989).

Another type of fimbriae, termed S-fimbria or S fimbrial adhesin (Sfa; Moch *et al.*, 1987; Schmoll *et al.*, 1989), binds to sialic acid-containing receptors (Korhonen *et al.*, 1984). These fimbriae are specifically associated with *E. coli* strains that cause septicaemia and meningitis. These fimbriae are encoded by a cluster of genes with a similar organisation to the type I system (Schmoll *et al.*, 1989). Sfa are produced by strains exhibiting the K1 antigen, such as O18:K1 or O83:K1 isolates (Korhonen *et al.*, 1985).

Another new type of fimbria, expressed by an *E. coli* blood culture isolate of serotype O75:K1:H7, showed high degree of sequence homology to the Sfa and F1C (*F*imbriae of serotype One *C*) gene clusters (Van Die *et al.*, 1984b). This new type of fimbria is therefore termed S/F1C-related fimbriae (Sfr). Another fimbrial adhesin, found to be associated with O75:K5 *E. coli* strains, is the Dr haemagglutinin (previously termed the O75X fimbria-like adhesin) which was cloned recently and found to recognise a receptor on the Dr blood group antigen (Nowicki *et al.*, 1988).

Normark and colleagues have also independently cloned the second copy of the *pap*-homologous sequences from strain J96 and have designated this *prs* (*p*ap *r*elated

sequence or *pap*-2; Karr *et al.*, 1989; Lund *et al.*, 1987, 1988). The Prs pili are characterised by their ability to haemagglutinate sheep erythrocytes but not human erythrocytes, clearly exhibiting a different binding specificity from P pili (Lund *et al.*, 1988). However, Prs pili resemble P pili in several respects; the Prs pilus filament expressed by *E. coli* J96, is serologically indistinguishable from the P pilus filament expressed by the same strain, and the *prs* operon is homologous to the *pap* operon (Lund *et al.*, 1988). Subsequent studies have revealed that the only detectable difference between the *pap* and *prs* operons is limited to their respective PapG molecules (Stromberg *et al.*, 1990). These findings suggest that alterations in the receptor binding component of pili can result in significant changes in binding specificity. The subtype of P fimbriae, encoded by the *pap*-2 gene cluster, is capable of binding to Bowman's capsule in the human kidney (Karr *et al.*, 1989). The potential receptors for *pap*-2- (or *prs*-) encoded fimbriae in the human urinary tract are the antigen known as LKE (Luke) on human erythrocytes and the stage-specific embryonic antigen 4 (SSEA-4) in tissues (Karr *et al.*, 1990). Thus, pathogenic *E. coli*, such as J96, can express a variety of adhesive pili with distinct receptor binding properties that may be important in increasing the range of hosts and tissues that the organisms can colonise.

The genes encoding several serologically different P-fimbriae have been cloned and the genetic organisations of gene clusters involved in fimbrial biosynthesis have been studied (Clegg and Pierce, 1983; Normark *et al.*, 1983b; Van Die and Bergmans, 1984; Van Die *et al.*, 1984a, 1985). It has been shown that, in addition to the gene encoding the major subunit, at least eight other genes are involved in the expression of fimbriae and adhesion (Lindberg *et al.*, 1987). For the Pap (F13) P-fimbriae it has been shown that *papG* is the actual adhesin gene (Lindberg *et al.*, 1987), while *papE* and *papF* encode minor fimbrial subunits.

The gene clusters encoding the $F7_1$ and $F7_2$ fimbriae have been cloned from a uropathogenic *E. coli* strain (Van Die *et al.*, 1984a; 1985). The nomenclature *fso* (F

Seven One) and *fst* (F Seven Two) has been proposed for these gene clusters. The lettering A-G is used for the individual genes, in accordance with the Pap (F13) gene cluster nomenclature (Norgren *et al.*, 1984). Riegman *et al.* (1988) demonstrated that both FsoE and FstE proteins are non-adhesive minor fimbrial subunits, located at the tip of the fimbrial structure. The FsoF and FstF/G proteins play important roles in the initiation of polymerisation of the minor and major subunits into the fimbrial structure.

A new blood group-specific agglutinin has been identified on human pyelonephritogenic *E. coli* strains (Rhen *et al.*, 1986). This haemagglutin recognised the terminal N-acetyl-D-glucosamine, and was associated with a new type of fimbriae, G-fimbriae.

1.3.2.2 MRHA fimbriae of Enterotoxigenic *Escherichia coli*

MR fimbriae of enterotoxigenic *Escherichia coli* are serologically and morphologically heterogeneous (Paranchych and Frost, 1988). The first fimbriae identified (plasmid-encoded) were the heat-labile K88 antigen found in ETEC strains that cause diarrhoea in neonatal piglets.

Adhesion fimbriae tend to be rather species-specific. Thus, the K88 fimbriae are encountered only in strains pathogenic to piglets (Smith and Linggood, 1971). K99 (plasmid encoded) and F41 (chromosomally encoded) are associated with calf, lamb and piglet pathogens (Moon *et al.*, 1977; Orskov *et al.*, 1975), and the 987P fimbriae (chromosomally encoded) are found only in porcine strains (Isaacson *et al.*, 1977). K88 and F41 fimbriae are thinner (2-3 nm diameter) and more flexible than type I fimbriae and have been referred to by some workers as fibrillae. K99 fimbriae are of an intermediate size (about 4.5 nm in diameter).

Another fimbrial adhesin, was described in bovine strains associated with neonatal calf diarrhoea (Girardeau *et al.*, 1980; Pohl *et al.*, 1982). This fimbrial adhesin, which was classified as F17, mediates binding to receptors containing N-acetylglucosamine (NAG) on calf intestinal mucosal cells (Lintermans *et al.*, 1988).

Most likely, the genetic determinants for F17 fimbriae are located on the bacterial chromosome (Lintermans *et al.*, 1988). The fimbrial subunits are assembled from the F17-A peptides. Analysis of the F17 gene cluster indicated that the F17-A, F17-C, F17-D and F17-G genes are indispensable for the binding of the fimbriae to the intestinal villi. The gene for F17-G was localised and sequenced. F17-mediated binding requires the presence of an intact pilus structure, together with the binding moiety. The specific NAG-dependent adhesion is mediated by the F17-G product (Lintermans *et al.*, 1991).

A new K88-related antigen, designated as CS31A (plasmid-encoded), was isolated from diarrhoeic or septicaemic calves. CS31A consists of thin (2nm) "fibrillar" fimbriae (Girardeau *et al.*, 1988). The *clpG* (26.7 kDa) gene, encoding the CS31A subunit, was localised and the nucleotide sequence of the coding region was determined (Girardeau *et al.*, 1991).

Evans *et al.* (1975) first gave the name of Colonization Factor Antigen (CFA) to MR fimbriae on human ETEC. A similar fimbrial antigen that was found a few years later was designated as CFA/II (Evans and Evans, 1978) and the original antigen was designated CFA/I. CFA/II was subsequently found to consist of three distinct fimbriae that were termed Coli Surface antigens 1, 2 and 3 (CS1, CS2 and CS3; Cravioto *et al.*, 1979; 1982; Smyth, 1982). The terms Colonization Factor Antigens (CFAs), coli surface antigens (CS) and Putative Colonization Factor (PCF) are used interchangeably to describe these fimbriae. The name PCF was used until it was confirmed that they were CFAs. CS are subclasses of a particular CFA/PCF. The fimbriae associated with gastrointestinal disease in humans have been divided into at least four main CFAs: CFA/I (Evans *et al.*, 1975), CFA/II (Evans and Evans, 1978), CFA/III (Honda *et al.*, 1984) and CFA/IV (previously called PCF8775; Cravioto *et al.*, 1982; McConnell *et al.*, 1988; Thomas *et al.*, 1982, 1985), and two PCFs: PCFO159:H4 (Tacket *et al.*, 1987) and PCFO9 (Heuzenroeder *et al.*, 1990).

CFA/II is subdivided into CS1, CS2 and CS3, whereas CFA/IV (PCF8775) is

subdivided into CS4, CS5 and CS6. Some of these fimbrial types found on human ETEC appear to be present on only a relatively small number of serovars or serotypes. Thus, CS1 appear on strains of serotype O6: K15: H16 (or H-) and PCFO159:H4 strains of serotype O159:H4 or H20, whereas others occur naturally on strains belonging to a range of O-serovars, e.g. CFA/I, CS3, CS6 and CS5/CS6. The reason for this difference in distribution of fimbrial occurrence in wild type isolates is unclear. Moreover, ETEC expressing these different adhesins possess characteristic MR haemagglutination and enterotoxin-producing phenotypes. The different haemagglutination patterns presumably reflect different receptor specificities on the red blood cells for the different fimbrial ligands, but little is known about these receptors (Mouricout and Julien, 1986; Sjoberg *et al.*, 1988; Smyth, 1982).

1.4 Morphology and structure of fimbriae.

The first report of fimbriae in 1950 involved an electron microscopic study (Houwink and Van Itersen, 1950). The pili are distributed peritrichously on most Gram-negative bacteria, the number per cell dependent upon the organism and conditions of growth. On the other hand, the pili of myxobacteria are usually observed as tufts at the cell poles (MacRae *et al.*, 1977). Fimbriae have diameters in the range of 2-7nm. They are therefore considerably thinner than flagella, which have a diameter of ~20nm (Silverman and Simon, 1977), but equal to or thinner than F pili, whose diameter is 8nm (Brinton, 1965). Examples of the group of pili with morphologically similar thin, rigid, rod-like structures include *E. coli* type I pili, and fimbriae associated with urinary tract infections (Orskov *et al.*, 1980). S-fimbriae, 987P pili and F7 pili of *E. coli* probably consist of subunits packed into a three-dimensional array similar to the type I pili. The latter has axial holes (Brinton, 1965; Isaacson and Richter, 1981). These structures have a uniform cross sectional diameter of approximately 7nm and are composed of subunits with a molecular weight of 17,000. X-ray crystallography of the type I pilus revealed that it was arranged in a right-handed helix with a pitch distance of 6.32nm (3.125 subunits

per turn), with an axial pore of 2nm in diameter. The subunits are so arranged that the supramolecular arrangement gives a rigid rod-shaped structure. Some pili in this group are as thin as 2nm, but others with diameters greater than 10nm have also been reported (Brinton, 1965; Ottow, 1975).

The *E. coli* K99 pilus appears to be flexible, with the reported diameters ranging from 2 to 4.5 nm (Isaacson, 1977). K88 and F41 are visualised under the electron microscope as very thin, wavy threads with diameters of ~ 2nm (Gaastra and De Graaf, 1982; Rhen *et al.*, 1983). Morphologically, the fimbriae on human ETEC strains can be divided into those resembling type I fimbriae, which are thin, wiry, wavy or curly structures with a diameter of 2 to 4 nm resembling K88 and F41 fimbriae (e.g. CS3 fimbriae). It is thought that this latter type of fimbria is also comprised of structural subunits in a helical arrangement, but the difference in the number of subunits per turn of the helix results in an open-structured helix rather than a closed one around an axial pore (Jones and Isaacson, 1983). The pitch distances and numbers of subunits per turn of the helix depend on the molecular weights of the subunits. The open helical structure proposed for *Bordetella pertussis* fimbriae, with a pitch distance of 6.5nm and 2.5 repeating units per turn, may be used as a model for these flexible, wiry fimbriae of human ETEC (Steven *et al.*, 1986). Some strains of human ETEC express only a single fimbrial type - e.g. CFA/I, PCFO166 and PCFO159:(H4). Others occur more commonly in specific combinations, e.g. CFA/III fimbriae + CS6 antigen, CS4 fimbriae + CS6 antigen, CS5 fimbriae + CS6 antigen, and CS1 or CS2 fimbriae + CS3 fimbriae. The presence of more than one fimbrial gene cluster in a strain raises the possibility of interaction between gene clusters in the regulation of expression of genes, or in the biogenesis of fimbriae on the cell surface. This, however, has not yet been reported.

Physio-chemical variability among pilus structures makes the categorisation of the pili by morphology difficult, especially since no X-ray diffraction investigations on any of these *E. coli* adhesins have been described.

1.5 Structural aspects of the fimbriae.

Bacterial pili are polymers, composed of identical protein subunits called pilin (Jones and Isaacson, 1983). The number of subunits defines the molecular weight and length of an intact pilus. The forces holding pilin subunits together are unknown but do not include covalent bonds. These non-covalent forces apparently exert considerable strength, as the pili resist disruption by various denaturing agents, such as SDS, which readily disrupts most protein polymers (Eshdat *et al.*, 1981; Klemm, 1984; Klemm *et al.*, 1982; Salit *et al.*, 1983). It is reasonable to believe that hydrogen- and hydrophobic-bonds are important in pilus formation and maintenance. Two important functional constraints are obviously imposed on any fimbrial protein:

(1) it must supply the necessary interaction, hydrophobic or electrostatic, of subunits in order to conserve the structural integrity of the whole fimbria, and

(2) it must provide, by means of specific structural elements inherent in its amino acid sequence, a binding site for a specific host epithelial receptor in order to function as an adhesin.

Most fimbriae identified so far have now been purified and the molecular weights of their structural subunits characterised (Tables 1.1 and 1.2). The N-terminal amino acid sequences of several of these subunits have been determined using purified fimbriae and/or have been deduced from the nucleotide sequences of cloned genes (Tables 1.3 and 1.4). Each contains about 50% non-polar amino acids. The calculation of protein hydrophobicity, based on amino acid composition, results in a value that assumes a uniform distribution of hydrophobicity, which is unrealistic. Surface hydrophobicity has also been measured in a salting out procedure using ammonium sulphate (Lindahl *et al.*, 1981). The greater the amount of ammonium sulphate required to salt out the fimbrial structure, the lower its hydrophobicity. It is therefore possible to rank each pilus type according to its hydrophobicity. Although this procedure has been given limited use, the results are in close agreement with the calculated values (CFA/I > CFA/II > K88 >

Table 1.1

Properties of ETEC fimbriae.

Fimbriae	Morphology	MW	Reference
Type I	rigid	15,700	Orndorff and Falkow (1984)
K88	flexible	27,600	Mooi *et al.* (1979)
CS31A	flexible	29,500	Girardeau *et al.* (1988)
F41	flexible	29,500	Moseley *et al.* (1986)
CS1541	Flexible	18,000/19,000	Broes *et al.* (1988)
K99	flexible	16,500	Van Embden *et al.* (1980)
987P	rigid	17,100	De Graaf and Klaasen (1986)
F17	flexible	19,500	Girardeau *et al.* (1980)
CFA/I	rigid	15,000	Klemm (1982)
		13,000	Willshaw *et al.* (1985)
		13,000	Smyth (1982)
CFA/II:			
CS1	rigid	16,300	Smyth (1982)
			Hall *et al.* (1989)
CS2	rigid	17,000	Klemm *et al.* (1985a)
		16,000	Sjoberg *et al.* (1988)
		15,800	Smyth (1982)
CS3	wiry	16,000	Boylan *et al.* (1988b)
		14,800	Smyth (1982)
		14,500/15,500	Levine *et al.* (1984)
		16,600/18,000	Manning *et al.* (1985)
CFA/III	rigid	18,000	Honda *et al.* (1984)
CFA/IV:			
CS4	rigid	22,000	Wolf *et al.* (1989)
		17,000	McConnell *et al.* (1988)
CS5	rigid/wiry	23,000	Heuzenroeder *et al.* (1989)
		21,000	McConnell *et al.* (1988)
CS6	?	14,500;16,000	Wolf *et al.* (1989)
CS7	?	15,500;17,000	Hibberd *et al.* (1990)
CS17	rigid	17,500	McConnell *et al.* (1990)
PCFO148:H2	wiry	n.d	Knutton *et al.* (1987)
PCFO159:H4	rigid	19,000	Tacket *et al.* (1987)
PCFO166	rigid	15,500;17,000	McConnell *et al.* (1989)
PCFO9	flexible	27,000	Heuzenroeder *et al.* (1990)
Antigen 8786	wiry?	16,300	Aubel *et al.* (1991)
Antigen 2320	flexible	16,000	Darfeuille *et al.* (1986)

Table 1.2

Summary of human fimbrial polypeptides

Fimbriae	Gene	Size (kDa)	Putative function
CFA/I:			
Region 1	*cfaA*	23.5	Shuttle function in transport of subunits ?
	cfaB	15	Structural subunit
	cfaC	85	Anchor in outer membrane ?
		38.3	
		29.4	
		26.9	
Region 2		25.7	
		14.9	
		13.8	
	cfaD	30	Regulation
CFA/II:			
CS3		97	Anchor in outer membrane?
		94	
		58	
		46.5	
		26/27	
		24	
		15	Pilin subunit
CS1/CS2		30	Regulation (Rns)
CS1		16.3	Pilin subunit
CS2		17	Pilin subunit

Table 1.3

N-terminal amino acid sequences of fimbrial subunits

Fimbria	N-terminal amino acids	Reference
F11	APTIPQGQGKVTFNGTVVDAP	Van Die et. al. (1986)
F13 (PapA)	APTIPQGQGKVTFNGTVVDA	Båga *et al.* (1984)
$F7_1$	AASIPQGQGEVSFKGTVVDAP	Van Die *et al.* (1985)
$F7_2$	APTIPQGQGKVTFNGTVVDAP	Van Die *et al.* (1984a)
SfaA	VTT-VNGGTVHFKGEVVDAA	Schmoll *et al.* (1989)
FIC	VTT-VNGGTVHFKGEVVNAA	Van Die *et al.* (1984b)
PilA	AATT-VNGGTVHFKGEVVNAA	Orndorff and Falkow (1985)
FimA	AATT-VNGGTVHFKGEVVNAA	Klemm (1984)
Type I b	ATT-VNGGTVHFKGEVV	Klemm (1985)
Type I c	VTT-VNGGTVHFKGEVV	Rhen *et al.* (1985)
K99	NTTINFNGKITSAT	Roosendaal *et al.* (1984)
K88ab(p17.6)	AVQKTIFSADVVASV	Mooi *et al.* (1984)
K88ab (p26)	WMTGDFNGSVDIGG	Gaastra *et al.* (1981)
E. coli F7	AATIPQGGEVAFKGTVVDA	Klemm (1985)
K. pneumoniae	NTTTVNGGTVAFKGEVVDA	Fader *et al.* (1982)

Table 1.4

N-terminal amino acid sequences of fimbrial subunits in human ETEC

Fimbria	N-Terminal Amino Acids	Reference
CFA/I	VEKNITVTASVDPVIDLLQADGNALPSAVK	Hamers *et al.*(1989); Klemm (1982)
CS1	VEKTESVTASVDPTVDLLQ	Hall *et al.* (1989).
CS2	AEKNITVTASVDPVIDLLQA	Klemm *et al.* (1985a).
CS3	AAGPTLTKELALNVLSPAALDATDAT	Boylan *et al.* (1988b)
CS4	VELNITVCASVDPTICIKQA	Wolf *et al.* (1989)
CS5	AVTNGQLTFNWQGVVPSAPVTQSSQPFVNG	Clark *et al.* (1992)
PCFO9	DSQQDSAFNGNIELGGTLSPEVKKLPRELR	Heuzenroeder *et al.* (1990)
8786	AGFVGNKAVVQAAVTIAAQNTTS	Aubel *et al.* (1991)
2230	GNVLSGGNGTQTVTMPVNAATXTVSMPTDPTD	Darfeuille *et al.* (1986)

K99 > type I). It is probable that the hydrophobic domains at the receptor recognition site may promote the interaction with the receptor.

The large K88 fimbrial subunit was the first for which the amino acid sequence was determined. At least three serological variants have been described, designated K88ab, K88ac and K88ad (Guinee and Jansen, 1979; Orskov *et al.*, 1964). The differences in amino acid composition among the K88 variants suggest that the serological differences are solely due to differences in the primary structures of the proteins.

The N-terminal amino acid sequences of several pili show remarkable relatedness. The sequences of the first 20 aa of different type I pili of *E. coli* (Klemm *et al.*, 1982) and *Klebsiella pneumoniae* (Fader *et al.*, 1982) and that of the *E. coli* F7 pilus (Klemm *et al.*, 1982) are nearly identical. However, the five pili are antigenically unrelated (Table 1.3). Unlike the others, haemagglutination of erythrocytes by the F7 pilus is not inhibited by D-mannose (Fader *et al.*, 1982; Klemm *et al.*, 1982).

In the case of some fimbrial subunit proteins (e.g. those of KS71A, F72, Pap and F11 fimbriae; Van Die and Bergmans, 1984; Van Die *et al.*, 1984a; 1985) the conservation of amino acid residues is most apparent at the C-termini and to a lesser extent at the N-termini. These regions are probably involved in functions common to this group of proteins, such as transport across the outer membrane, anchorage and maintenance of the fimbrial structure. The more variable parts have probably been adapted for the recognition of different receptors, or may simply have diverged as a response to immunological pressures. The conserved N-termini and C-termini are more hydrophobic than the non-conserved regions of their primary structures (Van Die *et al.*, 1987). It has been suggested that the non-conserved regions are located at the outside of the subunits and also on the outer surface of the fimbriae (Van Die *et al.*, 1987).

N-terminal amino acid sequence homologies between *E. coli* K88 pili (Gaastra *et al.*, 1979, 1981, 1983; Klemm, 1981), K99 pili (Roosendaal *et al.*, 1984) and, to a lesser

extent, F41 pili (De Graaf and Roorda, 1982) have been observed. The type I, F7 to F12, Pap and K99 fimbriae form a distinct group, as suggested by their amino acid sequences (Klemm, 1985; De Graaf and Mooi, 1986). Likewise, the K88 fimbrial series of antigenic variants form a distinct fimbrial group. Among the rod-shaped fimbriae of human ETEC, there is a substantial degree of homology between the N-terminal amino acid sequences of CS1, CS2, CS4 and CFA/I fimbrial subunits which indicate a group of evolutionally related fimbriae (Table 1.4). In contrast, the N-terminal amino acid sequences of CS3 and CS5 fimbriae show no close homology, either to each other or to those of the CFA/I, CS1, CS2 or CS4. These fimbrial subunits therefore appear to belong to distinct groups.

Although the pilus structure does not contain inter-pilin covalent bonds, it is likely that intra-pilin covalent bonds are present in some pili. For example, pilin molecules that contain di-sulphide bonds may assume three different conformations, depending upon the integrity of the bond (Isaacson, 1980b; Isaacson and Richter, 1981; Isaacson *et al.*, 1981; Jann *et al.*, 1981; McMichael and Ou, 1979). Many, but not all pilin molecules contain disulphide bonds. The fact that each of them assumes three different conformations suggests the existence of conserved amino acids in their primary sequences.

Amino acid analyses of purified fimbriae have revealed a lack of Cys residues in CS1 (Hall *et al.*, 1989), CS2 (Klemm *et al.*, 1985a; Sjoberg *et al.*, 1988), CS3 (Boylan *et al.*, 1988b) and CFA/I (Klemm, 1982) fimbrial subunits. In contrast, CS4 fimbrial subunits have at least two Cys residues. Therefore, CFA/I, CS1, CS2 and CS3 fimbrial subunits must be held together by non-covalent bonds. These fimbrial subunits also contain few aromatic amino acid residues. This contrasts with K88 fimbriae. CS1, CS2, CFA/I and type I fimbrial subunits each contain only two tyrosine residues. These residues are at the C-terminus in the case of the type I fimbrial subunit. It is therefore of interest that aromatic amino acids have been implicated in the maintenance of the

quaternary structure of some fimbriae (Mooi and De Graaf, 1985; Watts *et al.*, 1983).

1.6 Biogenesis of fimbriae

The genetic organisation of several *E. coli* pilus adhesins has been determined by recombinant DNA technology. Although there are noteworthy differences in the genetic organisation of the known fimbrial operons, some basic constraints are imposed on such systems. For example, various auxiliary proteins are needed for the subunits to be processed, translocated and assembled into intact fimbriae. The type I and Pap pili operons encode several accessory proteins required for regulation and biogenesis of the pilus. Pap pili are morphologically similar to type I fimbriae. Generally five to eight polypeptides are implicated in the phenotypic expression of *E. coli* fimbriae (Båga *et al.*, 1985; De Graaf *et al.*, 1984; Klemm *et al.*, 1985b; Mooi *et al.*, 1982; Norgren *et al.*, 1984; Orndorff and Falkow, 1984; Van Die *et al.*, 1984a). The genes encoding the subunit and auxiliary proteins are normally placed close together and are in many cases found on a common transcriptional unit (De Graaf *et al.*, 1984; Dougan *et al.*, 1983; Mooi *et al.*, 1981).

1.6.1 Pap fimbrial system

The biogenesis of Gal-α–(1-4) Gal binding P-fimbriae requires the co-ordinate expression of at least eleven closely linked chromosomal *pap* genes (Båga *et al.*, 1985; Hultgren *et al.*, 1991). P pili, encoded by the *pap* operon, are composite fibers consisting of flexible adhesive fibrillae joined end-to-end to the pilus rod on uropathogenic *E. coli* (O'Hanley *et al.*, 1985; Kallenius *et al.*, 1981). The specialised fibrillar structures at the tips of P pili are composed of PapE, PapF and the adhesin PapG (Lindberg *et al.*, 1987, 1989; Lund *et al.*, 1987). Kuehn *et al.* (1992) reported that the adhesin is a component of distinct fibrillar structures present at the tips of the pili. These virulence-associated tip fibrillae are thin, flexible polymers composed mostly of repeating subunits of PapE that frequently terminate with the α-D-galactopyranosyl-(1-4)-β-D-galactopyranose- or

Gal-α(1-4) Gal-binding PapG adhesin (Lund *et al.*, 1987). The stalk of the pilus is composed of repeating PapA monomers, probably arranged in a right-handed helix (Brinton, 1965; Lindberg *et al.*, 1984; Uhlin *et al.*, 1985a). PapH, located at the base of the pilus, is a necessary component of the pilus anchor (Båga *et al.*, 1987). All of the genes encoding these pilus protein subunits have been sequenced and characterised (Båga *et al.*, 1984, 1987; Hultgren *et al.*, 1989, 1991; Lindberg *et al.*, 1986; Tennent *et al.*, 1990a). PapD binds to pilus subunit proteins after they are imported into the periplasmic space (Hultgren *et al.*, 1989, 1991; Lindberg *et al.*, 1989). Stable PapD-PapG and PapD-PapE periplasmic pre-assembly complexes have been purified (Hultgren *et al.*, 1989; Lindberg *et al.*, 1989). These stable complexes are transported to an assembly site, thought to be composed, in part, of the outer membrane protein PapC (Norgren *et al.*, 1987), where the complexes are dissociated and the pilus subunits are polymerised.

The mechanism by which the periplasmic PapD chaperone ensures correct interactions of six different types of pilus subunit proteins so that they are assembled into well defined composite fibers that have distinct adhesive tip structures is virtually unknown. Kuehn *et al.* (1991) proposed that the role of this protein is to bind to interactive assembly surfaces on its pilus protein targets to prevent non-productive aggregation of pilus subunits exported into the periplasm. The assembly of pili seems to involve the targeting of chaperone-pilus protein complexes to the outer membrane assembly sites, where PapD is dissociated by an unknown ATP-independent mechanism and the released interactive subunit is assembled into the pilus.

The *papJ* gene encodes an 18kDa polypeptide (Tennent *et al.*, 1990a), which may be involved in pilus assembly/chaperone integrity (Hultgren *et al.*, 1991). It too may function as a molecular chaperone, directly or indirectly establishing the correct assembly of PapA subunits in the P-pilus. The separation of adhesive from fimbrial structural determinants in the Pap system conceivably confers an advantage, in that it might be

possible for the bacterium to evade an immune response by using fimbriae with an alternative antigenic structure but the same binding specificity.

Studies on the transcriptional regulation of the Pap gene cluster have shown that the polycistronic transcript initiating with *papB* and including the major subunit gene, *papA*, and the monocistronic *papI* transcript are divergently oriented. Transcription of both occurs at 37°C but is turned off at 26°C. A positive *trans*-regulatory effect of PapB and PapI to the expression of the structural gene *papA* was also observed by Båga and co-workers (Båga *et al.*, 1985; Uhlin *et al.*, 1985b). The intergenic regions between the two regulatory genes and the structural gene are necessary in *cis*-orientation for the thermo-regulation of fimbrial production and they also play a role in the repressive effect of trimethoprim on P fimbrial expression (Vaisanen-Rhen *et al.*, 1988). In addition, a putative CRP (cAMP-receptor protein) binding site was detected in the intergenic region. This implies an influence by catabolite repression on the transcription of the *pap* operon (Båga *et al.*, 1985; Göransson and Uhlin, 1984). Båga *et al.* (1988) reported that the differences in the expression of the genes *papB*, *papA* and *papH*, which are all part of the same transcriptional unit, may result from different half-lives of the corresponding mRNA molecules. Regulation of the expression of the *pap* genes also involves post-transcriptional mechanisms. It has been shown that the *papB/papA* transcript is subject to specific endonucleolytic cleavage (Båga *et al.*, 1988). This possibility, together with the presence of a terminator downstream of *papA*, may explain why the amount of PapA protein produced is somewhat larger than those of PapB or PapH. Forsman *et al.* (1989) have shown that PapB has a dual function as an activator/represser of pilus adhesin transcription, and that its autoregulatory mode of action involves differential binding to specific DNA sequences located at different positions upstream and within its gene and promoter.

The fact that *E. coli* strains can contain multiple Pap-like (Forsman *et al.*, 1989) gene clusters raises the possibility that this mode of transcriptional control might be a

means whereby more than one kind of fimbriae, differing in adherent properties, can be differentially expressed.

Göransson *et al.* (1988) showed that a plasmid containing the *prs* genes is able to complement *papI/papB* double mutants and restore the production of functional Pap fimbriae. Similar results were obtained with *pap* genes derived from other isolates of uropathogenic *E. coli*, and with gene clusters specifying the production of fimbriae that adhere to sialic acid-containing receptors.

Phase variation refers to the reversible transition of bacteria between fimbriate and non-fimbriate states ($fim^+ \leftrightarrow fim^-$). Low *et al.* (1987) have found that transcription of the *papA* gene in *E. coli* is subjected to control by a heritable phase variation mechanism in which alternation between transcriptionally active (phase-on) and inactive (phase-off) states occur. Phase variation is responsive to at least two environmental signals: temperature and carbon source. Blyn *et al.* (1990) suggested that phase switching occurs without DNA rearrangement of *pap* DNA sequences, distinguishing this system from those described for *E. coli* type I pili and *Salmonella* flagellar phase variation. Analysis of the regulatory region upstream of *papA* in DNAs isolated from phase-on and phase-off cell populations showed that two deoxyadenosine methylase (Dam) sites, $GATC_{1028}$ and $GATC_{1130}$, were present. The presence of unmethylated GATC sites in *E. coli* is unusual and has not been previously reported. Blyn *et al.* (1989) suggest that the methylation states of $GATC_{1028}$ and $GATC_{1130}$ may affect the regulation of *pap* transcription. *papA* transcription was absent in *dam*$^-$ *E. coli*.

Using *papB Ap-lac* fusion experiments, White-Ziegler *et al.* (1990) identified a locus that was involved in the thermo-regulation of Pap pilin transcription and designated it *tcp* (*t*hermo-regulatory *c*ontrol of *pap*). The *tcp* locus mapped to approximately 23.4 min on the *E. coli* K-12 chromosome. It appeared to be distinct from the recently described *E. coli* homologue to the *S. flexneri virR* gene, which is located at about 27.5 min on the *E. coli* chromosome (Hromockyj and Maurelli, 1989). Thermo-regulation of

Pap pili has been shown to occur at the level of transcription (Göransson and Uhlin, 1984; Roosendaal *et al.*, 1986). The control of expression is brought about by temperature-dependent transcription of a regulatory cistron in the pilus adhesin gene cluster. Göransson *et al.* (1990) identified and characterised a histone-like bacterial protein that has an important role in the thermo-regulation of transcription, and the locus was denoted *drdX* (for *d*e*r*epressed e*x*pression). The *drdX* locus was mapped in the vicinity of the *trp* region at 27.5 min on the *E. coli* chromosome. The *drdX* gene is identical to a DNA-binding protein in *E. coli* called H-NS (Falconi *et al.*, 1988) and negatively regulates two divergently oriented promoters. The intercistronic region between *papI* and *papB* contains a divergent upstream activating sequence (UAS), which mediates the enhancing effect of cyclic AMP and of PapB on transcription of both promoters (Göransson *et al.*, 1989).

1.6.2 Type I fimbriae

The genetic organization of the type I pilus cluster is similar to the *pap* operon (Clegg *et al.*, 1985a, 1985b; Gerlach *et al.*, 1989; Klemm, *et al.*,1985b; Orndorff and Falkow, 1984; Tennent *et al.*, 1990b). Although fewer genes compose the type I gene cluster, their functions are extremely similar to corresponding genes in the *pap* operon, suggesting that the biosynthesis of these supramolecular structures is similar. The *E. coli* type I *fim* gene cluster includes, in addition to the structural gene *fimA*, two regulatory genes *fimB* and *fimE*, which directs the phase-dependent expression of the *fimA* gene (Klemm, 1986). In addition to *fimA* and its regulators, *fimC* and *fimD* have been identified and are necessary for the transport and assembly of pili (Klemm *et al.*, 1985b; Orndorff and Falkow, 1984). FimC is thought to be a periplasmic protein (Klemm *et al.*, 1985b; Orndorff and Falkow, 1984) and possibly functions in the translocation and assembly of the fimbrial proteins (Hultgren *et al.*, 1990) as suggested for PapD in the pap operon (Hultgren *et al* 1989; Lindberg *et al.*, 1989; Norgren *et al.*, 1984). FimC is also required for surface localization of the mannose-binding adhesin moiety even in the

absence of the major FimA subunit (Hultgren *et al.*, 1990). FimD resides in the outer membrane and might serve the same function as PapC in the assembly (Klemm and Christiansen, 1990). Three genes distal to the structural gene (*fimA*) have been identified, and their corresponding gene products appear to be minor components of the type I pilus (Abraham *et al.*, 1987; Hanson and Brinton, 1988; Klemm and Christiansen, 1987). The product of one of these genes, FimH, has been specifically implicated in mediating D-mannose-specific binding (Abraham *et al.*, 1987, 1988a, 1988b; Krogfelt *et al.*, 1990; Maurer and Orndorff, 1985, 1987). The protein FimF may be comparable to PapH in the Pap pili system. The latter is located at the base of the pilus and is believed to be the last pilus subunit incorporated into the pilus filament (Båga *et al.*, 1987; Hultgren *et al.*, 1991; Klemm and Christiansen, 1987; Maurer and Orndorff, 1987)

E. coli possessing type I fimbriae undergo phase variation due to the reversible switching of transcription of the major fimbrial subunit gene, *fimA* (Eisenstein, 1981). This involves the inversion of a 300-bp DNA segment harbouring the promoter for the *fimA* gene (Abraham *et al.*, 1985; Dorman and Higgins, 1987; Eisenstein, 1981). Abraham *et al.* (1985) have shown that the mechanism underlying this switch involves the inversion of a 314bp segment of DNA, which contains the *fimA* promoter, so that *fimA* is transcribed only in one orientation of this invertible region. The inversion is controlled by two trans-acting regulatory genes, *fimB* and *fimE*, which are located upstream to *fimA* (Klemm, 1986). The FimB protein switches the invertible *fimA* promoter to the 'on' orientation (*fimA* expressed), whilst FimE promotes the adoption of the 'off' orientation. In the *fimA* promoter region, Dorman and Higgins (1987) found two matches to the consensus sequence for IHF (Integration Host Factor) binding, and they showed that IHF indeed played a role in the transcription of *fimA*. Probably, this host protein may bind to the DNA and induce a conformational change necessary for the recombination, mediated by FimB and FimE proteins. Pallesen *et al.* (1989) suggested that the direction and frequency of switching is determined by the ratio and total

concentration of the two regulators. Possibly, once FimE has induced a change to the 'off' position it is then itself repressed by interfering with transcription from the *fimA* promoter, which in this configuration might produce an anti-sense *fimE* transcript.

It is not clear whether in the type I system the *fimA* promoter is subjected to any transcriptional control other than the promoter inversion mechanism. The *fimE* gene was reported by Orndorff and Falkow (1984) to be a repressor of *fimA* expression and not involved in the metastable promoter inversion, whereas Klemm (1986) found it to have some enhancing activity on *fimA* transcription.

1.6.3 F4 (K88) fimbriae

Four antigenically different fimbrial antigens, F4 (K88), F5 (K99), F6 (987P) and F41 are important virulence factors of porcine ETEC (Gaastra and de Graaf, 1982; Orskov and Orskov, 1983). The K88 antigen was the first fimbrial adhesin found to be plasmid encoded (Orskov and Orskov, 1966). It enables the ETEC strains to colonise the small intestine by specific adhesion to the mucosa.

So far, three subtypes of the F4 (K88) fimbrial antigen, F4ab, F4ac and F4ad (also designated as K88ab, K88ac and K88ad) have been described (Guinee and Jansen, 1979; Orskov *et al.*, 1964). These variants have very few differences in their amino acid sequences (Gaastra *et al.*, 1979, 1981, 1983; Mooi and De Graaf, 1979; Mooi *et al.*, 1984). The K88 antigen is composed of identical subunits with apparent molecular weights ranging from 23,500 to 26,000 in different variants (Gaastra *et al.*, 1981; Mooi and De Graaf, 1979). The organisation and expression of the genes involved in the production of K88 antigen have been described (Mooi *et al.*, 1981). The cloned DNA fragment responsible for the biosynthesis of K88 fimbriae has been shown to contain six genes designated *faeC*, *faeD*, *faeE*, *faeF*, *faeG*, and *faeH* (*fae* stands for *f*imbrial *a*dhesin *e*ighty-eight; Mooi and De Graaf 1979; Mooi *et al.*, 1981; De Graaf, 1990). The major subunit, encoded by *faeG*, itself is responsible for receptor binding (Jacobs *et al.*, 1987a). In addition to *faeG*, the gene cluster also codes for several minor components. FaeC, for

instance, is present at the tip of K88 fimbriae (Oudega *et al.*, 1989). At least two additional proteins, encoded by *faeE* and *faeD*, are required for translocation of fimbrial subunits across the periplasm and the outer membrane (Mooi *et al.*, 1982, 1983). The nucleotide sequence of the *faeD* gene has been determined and shown to encode a large outer membrane protein of 82 kDa (Mooi *et al.*, 1986; Van Doorn *et al.*, 1982). It is required for the translocation of fimbrial subunits across the outer membrane and for the assembly of subunits into a multimeric structure. It also plays a role in anchoring the fimbriae to the cell surface. The product of the *faeE* gene is located in the periplasm and has a size of 27 kDa (Van Doorn *et al.*, 1982; Mooi *et al.*, 1982). Bakker *et al.* (1991) have shown that FaeE functions as a chaperone-like protein. Its interactions with the fimbrial subunit (FaeG) in the periplasm to stabilise the latter and prevent its degradation by the cell-envelope protease. Furthermore, FaeE prevents the formation of FaeG multimers which cannot be incorporated into the fimbriae. Bakker *et al.* (1992) showed that several amino acid residues in FaeG were involved in the formation of the receptor binding site. A clear correlation was found between the receptor binding site and the serotype-specific antigenic determinants.

Gaastra and Pedersen (1985), Thiry *et al.* (1989) and Pedersen (1991) showed that, unlike the conserved regions, the variable regions among the serological K88 variants are not essential for their secretion and assembly. This was in agreement with results in other fimbrial systems such as type I fimbriae (Hedegaard and Klemm, 1989), and P-fimbriae (Van Die *et al.*, 1989).

Mooi and De Graaf (1985) suggested that gene expression might be subjected to regulation both at transcriptional and post-translational levels in K88 and K99 fimbrial biogenesis. The ancillary fimbrial proteins (*i.e.* the basal protein, assembly proteins and minor subunits) are required at lower levels than the subunit itself. They suggested that an appropriate ratio of fimbrial components might be maintained. De Graaf and co-workers proposed that once transcribed, the outer membrane proteins in both K88

(*faeD*) and K99 (*fanD*) are probably translated in a temperature-dependent manner. The initiation codons of these genes lie within regions of dyad symmetry, and might be prevented from interacting with ribosomal RNA by the formation of a stem-loop structure in the mRNA. This would provide a means of thermo-regulation of fimbrial expression, since translation could proceed efficiently only at temperatures high enough to destabilise the mRNA hairpin. This result is also consistent with the observation that K99 fimbrial production is repressed at temperatures below 30°C (De Graaf *et al.*, 1980). If the availability of the ancillary fimbrial proteins is the limiting factor in the formation of fimbriae, then the control of the concentration of these factors might be the most sensitive level at which regulation of fimbrial biogenesis could be effected by the cell. Unlike the type I pili, K88 does not appear to be subject to phase variation.

1.6.4 K99 fimbriae

The K99 (F5) antigen was the second fimbrial adhesin discovered to be plasmid encoded (Smith and Linggood, 1972). K99 is fibrillar in nature and is often found on non-invasive ETEC strains that cause neo-natal diarrhoea in calves, lambs and piglets (Gaastra and De Graaf, 1982). The fimbriae are mainly composed of a protein subunit (FanC; *f*imbrial *a*dhesin *n*inety-nine) with a size of 16.5 kDa, and are implicated in the binding of the bacteria to intestinal epithelial cells and some species of erythrocytes (Gaastra and De Graaf, 1982). The amino acid sequence of the K99 subunit has been deduced from the nucleotide sequence of its corresponding gene (Roosendaal *et al.*, 1984). A ganglioside glycolipid, identified as Neu5Gc-α–(2->3)-Galp-β–(1->4)-Glcp-β–(1->1)-Ceramide, functions as the K99 receptor on horse erythrocytes (Smit *et al.*, 1984). This moiety has also been in the mucosal scrapings of pig and calf small intestines (Teneberg *et al.*, 1990).

At least eight polypeptides are involved in the biosynthesis of K99 fibrillae. Two polypeptides, FanA (11 kDa) and FanB (11.7 kDa), function in the regulation of expression of the gene encoding the fibrillar subunit, FanC, and of five other structural

genes encoding FanD-H, respectively (Roosendaal *et al.*, 1989). FanC is not only the main component of the K99 fibrillae, but is also required for the interaction with the glycolipid receptor. Jacobs *et al.* (1987b) showed that a lysine residue at position 132 and an arginine at position 136 in FanC are essential for receptor binding, and probably constitute a part of the binding domain. FanD (84.5 kDa) is an outer membrane protein involved in export, assembly and/or anchorage of the fibrillar components (Roosendaal and De Graaf, 1989; Oudega and De Graaf, 1988). FanE (23 kDa) is a periplasmic protein which functions as carrier protein in the transport of fibrillar subunits from the cytoplasmic membrane to the outer membrane. It also protects FanC, FanF, FanG and FanH against proteolytic degradation (Bakker *et al.*, 1991). FanG (16.5 kDa) and FanH (16.7 kDa) are identified as fibrillar subunit-like polypeptides and the genes encoding these polypeptides have been sequenced (Roosendaal *et al.*, 1987). FanF (33.9 kDa) is a minor component of K99 fibrillae, present at the tip and in, or along the shaft of, the fibrillar structures. The amino acid sequence of FanF showed similarity with the K88ab major subunit, FaeG. FanF is reported to have a role in initiation and elongation of K99 fibrillae formation. Experiments involving the purified receptor have shown that FanF is not required for binding of K99 fibrillae to the ganglioside receptor (Simons *et al.*, 1990).

In the K99 fimbrial operon, immediately downstream of the major subunit gene *fanC*, there is a sequence exhibiting dyad symmetry which probably acts as a transcriptional termination signal. Since the next gene (*fanD*) starts very close to *fanC*, it was proposed that *fanD*, together with other downstream genes, is co-transcribed with the *fanC* terminator. Once transcribed, it is possibly translated in a temperature-dependent manner. Although a similar genetic arrangement is present in K88, the gene *faeC*, which is equivalent to *fanC*, encodes a minor subunit (Oudega *et al.*, 1989). K99 does not appear to be subject to phase variation. Environmental factors known to influence K99 production include the level of aeration during growth,

the presence of glucose, L-alanine and cAMP and temperature (Brinton, 1965; De Graaf *et al.* 1980; Guinee *et al.*, 1977; Isaacson, 1980a; Isaacson and Richter, 1981).

1.7 Fimbrial biogenesis in Human ETEC

1.7.1 General aspects

Research on the genetics of fimbriae production in human ETEC is not as advanced as in other systems, such as the K88, type I and the P-fimbriae. The fimbriae in human ETEC are heterogeneous, both morphologically and antigenically, but they share some common characteristics. The genes encoding these fimbriae generally reside on high molecular weight plasmids (Gaastra and De Graaf, 1982) which also often encode toxin molecules. The fimbriae can generally mediate haemagglutination and their production is dependent on the growth medium and temperature.

Although there are differences in the organisation of the genetic elements involved in fimbrial biogenesis, the data available for the human ETEC fimbriae reveal features in common with the plasmid-encoded K88 and K99 (Mooi *et al.*, 1986) as well as the chromosomally encoded fimbriae such as type I and Pap fimbriae (Normark *et al.*, 1986; Van Die *et al.*, 1987). The operons generally consist of approximately eight genes. Accessory proteins of similar molecular weights to the gene products characterised for the K88 and Pap fimbrial operons have been identified among the gene products of the determinants for CFA/I, CS3 and CS5 fimbriae (Boylan *et al.*, 1987; Clark *et al.*, 1992; Gaastra *et al.*, 1990; Heuzenroeder *et al.*, 1989; Manning *et al.*, 1985; Savelkoul *et al.*, 1990; Willshaw *et al.*, 1985).

Genes, or gene products, associated with the expression of fimbriae on human ETEC strains are summarised in Tables 1.2. Putative functions are assigned to these on the basis of knowledge gained from the K88, K99 and Pap fimbrial systems (Table 1.2).

1.7.2 Colonization Factor Antigen I (CFA/I)

An ETEC fimbrial adhesin that conferred specific adherence to human intestinal epithelium was first reported by Evans *et al.* (1975) and was designated colonization

factor antigen I. The CFA/I adhesin is encoded by non-conjugative plasmids of 52-65 MDa in size which often code for enterotoxins (Evans *et al.*, 1975). Unlike other fimbrial operons, the genetic information for the expression and assembly of CFA/I fimbriae is located in two regions, separated by about 40 kb (Smith *et al.*, 1982). These have been studied in detail and designated regions 1 and 2 (Smith *et al.*, 1982; Willshaw *et al.*, 1983). Region 1 directs the synthesis of at least six polypeptides, of which one is the major CFA/I subunit. Three other polypeptides appear to be synthesised as precursors which undergo processing (Willshaw *et al.*, 1985). Region 2 encodes three polypeptides, of which one is essential for the assembly of the fimbriae (Willshaw *et al.*, 1985; Gaastra *et al.*, 1990).

The structural gene for the fimbrial subunit, *cfaB*, has been isolated and sequenced (Hamers *et al.*, 1989; Karjalainen *et al.*, 1989). The protein has an apparent size of 13.3 kDa for the mature protein and 15.6 kDa for the precursor. The deduced polypeptide sequence is composed of 170 amino acids, of which the first 23 amino acids are considered to be the signal peptide. The CFA/I gene has a typical ribosome binding site located 10 base pairs (bp) upstream from the initiation codon, and a putative promoter at 48 bp upstream from the initiation codon. A pair of inverted repeats, followed by a stretch of eight adenine residues, is located 45 bp downstream from the termination codon of the CFA/I gene. This region may be a Rho-independent transcriptional terminator.

Upstream to *cfaB* is the gene *cfaA*, which encodes a protein of 24 kDa, while downstream is the gene *cfaC*, probably encoding an 85 kDa protein. The region in front of the *cfaA* gene contains a sequence homologous with the left terminal repeat of the IS2 insertion element (Ghosal *et al.*, 1979).

The nucleotide sequence of region 2 displays an open reading frame, designated as *cfaD* by Savelkoul *et al.* (1990) and *cfaR* by Caron and Scott (1990), encoding a protein of 265 amino acids. Its product has no signal sequence, and computer analysis

suggests it to be a DNA-binding protein related to the AraC family of transcriptional activators. It has been shown that *cfaD* enhances expression from the promoter of region 1. The gene, *cfaR*, is homologous to *rns* (for *r*egulatio*n* of CS1 and 2 antigens of CFA/II; Caron *et al.*, 1989), a positive regulator of the region 1 of CFA/I genes (Caron and Scott, 1990). Although *cfaR*, or *cfaD*, differs from *rns* at 28 positions, 14 of which alter the amino acid sequence (Caron *et al.*, 1989), they can functionally substitute for one another (Savelkoul *et al.*, 1990).

Gaastra *et al.* (1990) identified a DNA sequence, homologous to the *cfaD* gene of CFA/I region 2, on CFA/I region 1. This putative gene is designated *cfaD'*. It differs from the *cfaD* gene in containing two deletions and a stop codon. The *cfa*D' sequence therefore can only encode a truncated CfaD-like protein. A regulatory function for *cfaD'* is not likely since its deletion does not affect the production of CFA/I fimbriae in *E. coli* K-12 strains. However, *cfaD'* is present on wild-type plasmids isolated from CFA/I strains of different serotypes, obtained at various geographical locations.This suggests that it is not completely without a function.

The CFA/I pilin protein contains neither any cysteine residue nor substituted amino sugars (Klemm, 1982). It contains a rather hydrophobic C-terminal region and it has been proposed that this part of the molecule helps to maintain the integrity of the fimbrial super-structure by taking part in interactions between subunits, like the K88 protein (Klemm, 1982). The amino acid sequence of CFA/I does not show any homology with the known sequences of other fimbriae. On the basis of primary structure, the localisation of potential antigenic determinants of the CFA/I adhesin has been predicted (Klemm and Mikkelsen, 1982). Pieroni *et al.* (1988) identified a 26 kDa sialoglycoprotein, as a receptor for the CFA/I pili in human erythrocyte membranes. The specific binding components have been identified (Wenneras *et al.*, 1990).

CFA/I does not appear to be subject to phase variation. Karjalainen *et al.* (1991) indicated that the expression of CFA/I fimbriae was responsive to fluctuations in iron

concentration and showed that the repressive effect of iron on fimbrial production may be mediated by the *fur* gene product (*fur* stands for *f*erric *u*ptake *r*egulation) *via* the promoter for *cfaB*.

1.7.3 Colonization Factor Antigen II (CFA/II)

The second adhesive factor, CFA/II, was originally described as a single fimbrial antigen which caused MRHA of bovine erythrocytes (Evans and Evans, 1978). It is now realised that CFA/II is composed of more than one antigenic component. Using putative antiserum to CFA/II, Smyth (1982) identified and described three components called coli surface antigens CS1, CS2 and CS3. These antigens were associated with protein bands of 16.3, 15.3 and 14.8 kDa, respectively, in SDS gel electrophoresis. Smyth found a correlation between the presence of CS1 and CS2 and the haemagglutination of bovine erythrocytes. In contrast, CS3 showed no such activity. The three antigens described by Smyth appeared to be identical to component 1, 2 and 3 previously reported by Cravioto *et al.* (1982).

Nearly all CFA/II positive *E. coli* possess CS3, which probably corresponds to one of the components in the original CFA/II defined by Evans and Evans (1978). Furthermore, a strong correlation was found between CS fimbria and the presence of heat-labile and heat-stable enterotoxins. The CS fimbrial phenotype of an ETEC strain is dependent on the specific serotype and biotype of the host bacterium, which are therefore useful phenotypic markers of predictive value for fimbrial expression (Smyth, 1986). With the exception of one wild type ETEC of serotype O139:H28 (Scotland *et al.*, 1985), only ETEC of serotype O6: K15: H16 or H$^-$ belonging to biotype A (Scotland *et al.*, 1977) produce CS1, whereas biotypes B, C and F of the same serotype produce CS2 fimbriae. All biotypes produce CS3.

Mobilisation of plasmids encoding CS-fimbriae from serotype O6:H15:K16 or H$^-$ background to other O6 serotypes, to other O-serovars or to *E. coli* K-12 gives rise to expression of only the CS3 fimbriae in these new hosts. In contrast, depending upon the

biotype, the reverse mobilisation of plasmids from a CS3-only wild type background to serotype O6: K15:H16 or H$^-$ strains lacking a CS fimbriae-associated plasmid gives rise to CS1 or CS2 fimbriae (Smyth, 1986). A K15 or H16 phenotype, in combination with other O-serovars, does not allow expression of CS1 or CS2 fimbriae in such *E. coli* hosts. Twohig *et al.* (1988) indicated that expression of CS1 or CS2 fimbriae probably involves chromosomal determinants that are found only in strains of serotype O6:K15:H16 or H$^-$. All other natural ETEC hosts of such CS-fimbriae-associated plasmids produce only CS3 fimbriae, irrespective of O-serovar or O: K: H serotype and biotype. Smyth (1986) suggests that the expression of CS1 and CS2 fimbriae requires genetic information from the host chromosome as well as from the plasmid.

The morphological characteristics of CS1 and CS2 were shown to be much like those of the type I and CFA/I group. These antigens were visualised as stiff rod-like fimbriae with diameters of 6-7nm (Levine *et al.*, 1984; Mullany *et al.*, 1983; Smyth, 1982). In contrast, the CS3 antigen appeared to resemble the K88 and F41 antigens as thin, flexible fibrillar fimbriae, with diameters close to 2nm (Knutton *et al.*, 1984; Levine *et al.*, 1984).

The specific binding components for the different sub-components of CFA/II (CS1, CS2 and CS3) have been identified as glycoproteins (Wenneras *et al.*, 1990). Using immuno-thin layer chromatography, Oro *et al.* (1990) found that CFA/II and its three sub-components bound to a sialoganglioside GM1. Sjoberg *et al.* (1988) have shown the binding specificity of CS2 fimbriae for sialic acids.

1.7.3.1 CS1 and CS2

Molecular cloning of plasmid-encoded genetic determinants for the expression of CS1 and CS2 has been reported (Boylan *et al.*, 1988a; Willshaw *et al.*, 1988). Studies with two similar non-conjugative plasmids of 59 MDa, both of which were capable of producing CS1, CS2 and CS3, found that the expression of these fimbrial adhesins from a given plasmid was dependent on the biotype of the host strain (Mullany *et al.*, 1983).

More specifically, depending on the biotype of the cell, CS1 and CS3 together, CS2 and CS3 together or CS3 alone were produced. The cloning of CS1/CS2 fimbrial determinants demonstrated that the expression of CS1 and CS2 fimbriae parallelled the host-related restriction of expression observed with the parental plasmids, in *E. coli* K-12 and serotype O6:K15:H16 or H⁻ ETEC strains of different biotypes (Boylan and Smyth, 1985; Boylan *et al.*, 1988a; Willshaw *et al.*, 1988).

Boylan *et al.* (1988a) cloned the pili determinants from an ETEC strain, E90a, and expressed CS1, or CS2 and CS3, fimbrial adhesins in appropriate *E. coli* hosts. To characterise this further, a chimaeric plasmid called pCS200, which mediated the expression of CS1 or CS2, but not of CS3, fimbrial antigens in appropriate host strains was obtained. As previously observed with *E. coli* K-12, recipients of CS-fimbriae-associated plasmids mobilised from wild-type ETEC, the *E. coli* K-12 recipients of the pCS200 did not express the CS1 or CS2 fimbrial antigen. An oligonucleotide probe, synthesised on the basis of the published N-terminal amino acid sequence of the CS2 fimbrial subunit (Klemm *et al.*, 1985a), hybridised to plasmid pCS200, implying that the gene for the structural subunit resided on this plasmid. However, subsequent studies with pCS200 (Caron *et al.*, 1989) have found that this is not the case. There was just unfortunate homology of the redundant oligonucleotide sequence to one of the plasmid-encoded genes, called *rns* (*r*egulation of C*S*1 and 2) and required for the expression of CS1 and CS2. This gene encodes a regulatory protein that may act positively on the structural genes for these pili.

Production of CS1 by *E. coli* strain E24377, of serotype O139:H28, is controlled by a plasmid that encodes heat-stable and heat-labile enterotoxins as well as CS3. Willshaw *et al.* (1990a) have shown the presence, on this plasmid, of a sequence that regulates the expression of CFA/I fimbriae and is at least 96% homologous to the *rns* gene. A separate plasmid, encoding the structural genes for CS1 synthesis, was identified and transformed into the *E. coli* strain HB101 or a derivative of strain E24377 not

carrying any large plasmid. These transformants did not produce CS1 fimbrial antigen, but its expression was obtained when a cloned *cfaD* gene, or a wild-type plasmid carrying the *rns* gene, was introduced into the cell (Caron and Scott, 1990; *see* section 1.7.3.2). Transposon mutagenesis identified a 3.7 kb region which was essential for the production of CS1 fimbriae. Subsequently, all genes essential for the production of CS1 fimbriae were cloned on a 9.9 kb *Bam*HI fragment and expressed in the presence of *cfaD*.

The gene which encodes the major CS1 subunit, *coo*A (for coli surface antigen one), has been cloned from a plasmid different from the one encoding its regulator, the *rns* gene (Perez-casal *et al.*, 1990). The molecular mass calculated for the CooA protein is 15.2 kDa, in close agreement with the size observed on SDS-PAGE (16 kDa). The predicted amino acid content is consistent with that obtained by Hall *et al.* (1989) from the analysis of CS1 protein, but differs from that obtained by Smyth (1982). This difference probably reflects the different strains used. The analysis of the first gene in the *coo* locus, *cooB*, showed that its product, with an apparent molecular mass of 26 kDa, was required for assembly but not transport of the major pilin subunit, CooA (Scott *et al.*, 1992). There appear to be several similarities between the organisation of the genes involved in the expression of CFA/I and CS1, which suggest an evolutionary relationship between the operons involved. At the protein level, there is 92% similarity (55% identity) between CooA and CfaB, the major subunit of CFA/I pilin.

The CS2 fimbriae, with a subunit size of 17 kDa (Klemm *et al.*, 1985a; Table 1.1), preferentially agglutinate bovine erythrocytes (Faris *et al.*, 1982). The N-terminal amino acid sequence of CS2 is given in Table 1.4. This protein does not contain any cysteine residues (Klemm *et al.*, 1985a; Sjoberg *et al.*, 1988). Honda *et al.* (1989) reported that CS2-possessing ETEC were predominant among isolates of travellers' diarrhoea in Osaka, Japan. The pili from one clinical strain (TH61) were purified to homogeneity and found to be CS2 pili, with a subunit molecular weight of about 16,000.

1.7.3.2 Rns protein

The plasmid-encoded regulatory gene called *rns* is required for the expression of CS1 or CS2 colonization factor antigens. The CFA/II *rns* gene produces a 26 kDa protein with a pI of 10.1. The Rns protein shows significant amino acid sequence homology to the carboxy-terminal ends of the AraC, RhaR and RhaS proteins of *E. coli* and to the VirF proteins of both *Yersinia enterocolitica* and *Shigella flexneri.* AraC is the regulatory protein of the arabinose operon in *E. coli* (Caron *et al.*, 1989). In *Y. enterocolitica* VirF controls the virulence regulon associated with calcium dependency, secretion of outer-membrane proteins and pathogenicity (Cornelis *et al.*, 1989), whereas in *S. flexneri* VirF controls the expression of genes required for initial invasion of epithelial cells and for subsequent cell to cell spread (Sakai *et al.*, 1986, 1987).

The amino acid sequence homology between the Rns and AraC proteins includes the DNA-binding domain of the latter (aa 195-213; Caron *et al.*, 1989). In addition, the VirF proteins of *Y. enterocolitica* and *S. flexneri* have been shown to be transcriptional activators (Cornelis *et al.*, 1989; Sakai *et al.*, 1988) and, at least in *Y. enterocolitica*, VirF has been shown to regulate gene expression via DNA-binding (Cornelis *et al.*, 1989).

The average G+C content of *E. coli* DNA is 50%, yet the *rns* gene contains only 28% G+C, implying that it has been acquired from some other organism. Rns is 92% identical to the regulatory protein in the CFA/I system, CfaR or CfaD (Caron and Scott, 1990) and, like *rns,* the G+C content of *cfaR* is 28%. The *rns* gene also uses rare codons such as AGA/G for arginine, AUA for isoleucine and GGA/G for glycine (Caron *et al.*, 1989). Rare codons are generally used in genes which are not highly expressed, consistent with the proposed function of Rns as a regulatory protein. Different *rns* genes may be involved in the expression of CS1 and CS2 fimbriae. Boylan cloned one such gene in a plasmid (pCS300) which mediated the expression of CS2, but not of CS1, fimbriae (Smyth, personal comunication). The insert DNA in plasmid pCS200, from

which Caron *et al.* (1989) identified *rns*, differs in structure from the insert in plasmid pCS300.

The plasmid pEU2030 allowed the expression of CS1 or CS2 fimbrial antigens with subunit molecular weights corresponding to those reported for native CS1 and CS2 fimbriae (Table 1.2). Moreover, it produced MRHA of bovine erythrocytes. This can be interpreted to mean that it encoded the genes for the structural subunits of CS1 and CS2 fimbriae. The genes for accessory proteins required for the biogenesis must be chromosomally encoded. This explanation is consistent with the model for CS1 and CS2 fimbrial expression, proposed by Smyth (1986), to explain the host restricted nature of CS1 and CS2 fimbrial expression. The chromosomal location of the pilin genes explains the serotype-restricted production of CS1 and CS2 in the presence of the plasmid regulatory gene. The nucleotide sequence analysis by Caron *et al.* (1989), using the plasmid pEU2030, also provides an explanation for the DNA hybridisation data of Boylan *et al.* (1988a). The sequence demonstrates that the homology to the putative pilin gene probe was present upstream of the ATG start codon of the *rns* gene.

1.7.3.3 CS3

Compared to the other colonization factor antigens described above, relatively little is known about CS3. A family of recombinant plasmids encoding the genes for its synthesis has been cloned by Manning *et al.* (1985) from the wild type ETEC strain, PB176 (O6: H16 biotype A), the original CFA/II-expressing strain of Evans and Evans (1978). One of these plasmids, pPM474, carried a 4.7 kb DNA fragment and was used for further studies. A minimum coding region of at least 3.75 kb DNA was needed for the expression of CS3 fimbriae. Protein analysis identified seven polypeptides of 97, 94, 58, 46.5, 27, 18 and 16.6 kDa in sizes. Boylan *et al.* (1987) showed that the 15 kDa polypeptide was the structural subunit of CS3 fimbriae, and confirmed that the 17 kDa polypeptide was its precursor form. Since the total nucleotide sequence required to encode all these proteins exceeds 4.7 kb DNA, an overlapping arrangement of genes has

been implicated.

Boylan *et al.* (1987) have independently cloned a 5.1 kb *Hin*dIII DNA fragment from the wild-type plasmid (originally mobilised from the ETEC strain E90a of O6: H16: K15) which encodes the genetic determinants for CS3 synthesis in this strain. Subcloning and Tn*5* insertion mutagenesis of this CS3-expressing plasmid, pCS100, revealed that more than one cistron was involved in its expression. Polypeptides of 94, 26, 24, 17 and 15 kDa were detected in *E. coli* minicells.

The restriction maps of the plasmids, pPM474 and pCS100, appear to differ slightly (Manning *et al.*, 1985). This probably reflects the strain differences. The plasmids encoding CS3-fimbriae in PB176 and E90a are 60 and 51 MDa, respectively (Boylan and Smyth, 1985; Penaranda *et al.*, 1983), which further supports the assumption that strain variations do occur. As mentioned earlier, the plasmids associated with the production of CS1 and CS2 fimbriae have also been shown to be heterogeneous (Boylan and Smyth, 1985; Echeverria *et al.*, 1986; Mullany *et al.*, 1983; Penaranda *et al.*, 1983; Smith *et al.*, 1983).

While the work for this thesis was in progress, Boylan *et al.* (1988b) sequenced a 612 bp DNA fragment containing the gene for the major fimbrial subunit of CS3. A single open reading frame of 168 amino acids, with a calculated molecular mass of 18.4 kDa, was encoded by the sequence. A possible promoter region and a putative ribosome binding site were identified upstream of this open reading frame. A potential signal peptide cleavage site was identified between amino acid residues 21 and 22(Ala-Ala). This yields a processed subunit of 16 kDa which is slightly higher than the apparent size determined by SDS polyacrylamide gel electrophoresis (Table 1.1).

Willshaw *et al.* (1988) sub-cloned the plasmid DNA responsible for CS3 production in strain E5470 (serotype O6: H16, biotype C). They too found that a 5.4 kb fragment was all that was required for CS3 expression. The restriction map of this fragment showed similarity to that of pCS100 described earlier.

In summary, molecular cloning of the genetic determinants of CS3 fimbriae has been achieved independently by three groups (Manning *et al.*, 1985; Boylan *et al.*, 1987; Willshaw *et al.*, 1988). The cloned fragments carry a cluster of genes encoding a number of proteins necessary for the fimbrial biogenesis (Manning *et al.*, 1985; Boylan *et al.*, 1987). Regions associated with expression of specific gene products have been defined by deletion or transposon mutagenesis, coupled with analyses of proteins in minicell translations. All these analyses implied the presence of overlapping genes.

1.7.4 Colonization Factor Antigen III (CFA/III)

Honda *et al.* (1984) found a putative new colonization factor antigen of human ETEC and named it CFA/III. Darfeuille *et al.* (1983) also independently reported a CFA/III but Manning *et al.* (1985) suggested that the fimbriae described by these authors was in fact CFA/I. Under the electron microscope, CFA/III appear as rod-like fimbriae with a diameter of 6-8 nm (Honda *et al.*, 1984). Human ETEC colonization factors are restricted to a limited number of serogroups. The prototype CFA/III strain of Honda *et al.* (1984), which also produces CS6 (McConnell *et al.*, 1988), belongs to the serogroup O25:H$^-$.

1.7.4 Colonization Factor Antigen IV (CFA/IV)

A putative colonization factor, designated E8775, has been described for ETEC strains of serogroups O25, O115 and O167 (Thomas *et al.*, 1982) and was later renamed as putative colonization factor, PCF8775 (Thomas *et al.*, 1985). PCF8775 is analogous to CFA/II in that it is a complex composed of three antigens designated as CS4, CS5 and CS6. Later, McConnell *et al.* (1988) changed the name PCF8775 to Colonization Factor Antigen IV (CFA/IV) to conform with the nomenclature used for the other confirmed human colonization factors. Fimbrial antigens, CS4 and CS5, are rigid fimbriae with a diameter of about 6-7nm which promote MRHA of human and bovine erythrocytes. Two variants of the CS6 polypeptide, having sizes of 14.5 and 16.0 kDa, respectively, have been identified (McConnell *et al.*, 1988; Wolf *et al.*, 1989). These were thought to

represent two types of fimbriae, which differ in subunit sizes. CS6 is apparently non-fimbrial in nature, but the polypeptides might be assembled to form a polymeric structure as in the antigen 2230 (Darfeuille *et al.*, 1986). The DNA encoding CS6 has been cloned from a plasmid (Willshaw *et al.*, 1988). However, the structure of the third antigen, CS6 (Thomas *et al.*, 1985), remains uncertain; CS6 does not promote MRHA.

The fimbrial antigens of the CFA/IV complex (CS4, CS5 and CS6) are plasmid-encoded. The molecular characterisation of a new fimbrial type isolated from ETEC, belonging to the O115:H40 and O115:H$^-$ serotypes, was designated as CFA/V by Manning *et al.* (1986). This was later identified as the previously reported CS5 of PCF8775 (Heuzenroeder *et al.*, 1989). The sizes of CS4 and CS5 polypeptides are about 22.0 kDa and 23.0 kDa, respectively (Heuzenroeder *et al.*, 1989; Wolf *et al.*, 1989).

The N-terminal amino acid sequence of CS4 (Wolf *et al.*, 1989) was determined and found to be similar to that of CFA/I (Klemm, 1982) and CS2 (Klemm *et al.*, 1985a; Table 1.4). The N-terminal sequence of CS5 has shown it to be distinct from other fimbriae, although there is some conservation of residues (Heuzenroeder *et al.*, 1989). The genes determining the biosynthesis of CS5 fimbriae have been cloned and sequenced (Clark *et al.*, 1992; Heuzenroeder *et al.*, 1989). The fimbriae expressed from the cloned DNA appeared to differ morphologically from the CS5 fimbriae described by Thomas et al. (1985) in being wiry and 3nm in diameter, instead of rigid fimbriae of 6-7 nm in diameter. The wild type strains examined by Heuzenroeder *et al.* (1989) exhibited rigid rod-shaped fimbriae as well, possibly type I fimbriae, which did not react with antibodies directed against the 23 kDa CS5 fimbrial subunit protein.

Willshaw *et al.* (1990b) showed that two ETEC strains of serotype O25:H42, that produced CS4 and CS6, hybridised with a probe from the *cfaD* gene, which regulates expression of CFA/I. Transformation of a cloned *cfaD* gene, into derivatives of the strains that were negative for CS4 and CS6, resulted in expression of CS4 but not of CS6. By hybridisation, the sequence that regulated CS4 production in the wild type O25

strains was located on a plasmid that also encoded the CS6 antigen. The structural genes for the CS4 antigen were on a separate plasmid. The O25 strains carried a third plasmid encoding the genes for enterotoxin production, which were therefore not linked to the regulatory genes or genes encoding CS antigens. Willshaw *et al.* (1991) cloned the regulatory genes from CS5 fimbriae-producing strains of *E. coli* of O- serogroup O115 and O167:H5. The latter (O167 regulator) encodes a product sufficiently similar to those of *cfa*D, *rns* or CS4 regulator genes and promotes expression of CFA/I, CS1, CS2 and CS4 antigens. However, complete functional homology is lacking, as fimbriae are not always produced. Nucleotide sequence determination of this gene showed homology to *cfaD* (De Haan *et al.*, 1991).

"The attachment pili" (Deneke *et al.*, 1981), of *E. coli* strain 334A of serotype O15: H11, consist of subunits similar in size to CS5 subunits (21.5 kDa). The intact fimbriae of strain 334A have shown to be antigenically distinct from the other known colonization factors and have been designated CS7 (Hibberd *et al.*, 1990). Western immunoblotting revealed that the subunits of 334A and CS5 fimbriae shared common epitopes. Expression of 334A was controlled by a plasmid which also coded for enterotoxin production.

Thomas *et al.* (1985) noted the presence of curly, wiry fimbriae on some strains of $CS6^+$ bacteria. Such curly fimbriae have also been described on ETEC of serotype O148: H28 (Knutton *et al.*, 1987), a serovar known to produce CS6 antigen. These wiry fimbriae look morphologically similar to fimbriae recently described by Olsen *et al.* (1989), termed curli, as being fibronectin-binding fimbriae on *E. coli* strains associated with bovine mastitis. Curli are thin, wiry fibres with diameter of approximately 2nm and a size of 17 kDa. This differs from the known pilin protein and is synthesised in the absence of a cleavable signal peptide. The gene for the structural subunit, *crl*, is present and transcribed in most natural isolates of *E. coli*, but only certain strains are able to assemble the subunit protein into Curli pili. The assembly process occurs preferentially at

growth temperatures below 37°C.

Oro *et al.* (1990) examined the binding of colonization factor antigens of ETEC to gangliosides and asialo-gangliosides by using immuno-thin layer chromatography and found that the sub-component, CS4, of the CFA/IV complex bound to asialo ganglioside GM1.

1.7.6 PCFO9

The genes determining the biosynthesis of a new putative colonization factor, designated PCFO9, have been cloned from an ETEC strain of serotype O9: H$^-$ (Heuzenroeder *et al.*, 1990). A synthetic oligonucleotide probe, based on N-terminal amino acid sequence, has been used to define the region encoding the structural gene for the major pilin subunit (27 kDa). All the genes necessary for the biosynthesis of PCFO9 have been identified and sequenced (Clark, C.A., personal communication). The genetic organization and amino acid sequences of the products show a higher degree of homology to the porcine ETEC, K88.

1.7.7 Other human ETEC

Knutton *et al.* (1987) studied CFA in strains lacking MRHA properties and examined the ability of human ETEC strains to adhere to human proximal small intestinal mucosa. ETEC strain, B7A (O148:H28), was selected for study because it belonged to an epidemiologically important serotype and did not produce a known CFA. However, it is known to be pathogenic and has been shown to cause diarrhoeal disease in human volunteers. The results of a human enterocyte adhesion assay indicated that some bacteria in cultures of B7A produced a new fimbrial structure that was morphologically distinct from CFA/I, CFA/II and PCF8775 (CFA/IV) fimbriae. It consisted of curly fibrils of approximately 3 nm in diameter, which were readily identified when bacteria were subcultured from the mucosa and examined by electron microscopy. Identical fimbriae were produced by the ETEC strain 1782-77, of the same serotype. Identification of these fimbriae on only the bacteria subcultured from human intestinal mucosa strongly suggests

that they promote mucosal adhesion of ETEC serotype O148:H28 and thus represent a potentially new human CFA.

Tacket *et al.* (1987) found distinct plasmid-encoded fimbriae composed of 19 kDa protein subunits associated with the ETEC serotype O159:H4. This putative colonization factor of O159:H4 joins other ETEC fimbriae as potentially useful immunogens against human diarrhoea.

McConnell *et al.* (1989) described an ETEC strain of serogroup O166 which gave MRHA with bovine and human erythrocytes and found that it produced fimbriae of about 7 nm in diameter. Two possible fimbrial polypeptides, of molecular masses of 15.5 and 17.0 kDa, were seen on SDS-PAGE. The term putative colonization factor O166 (PCFO166) is proposed to describe the adhesive factor(s) on ETEC of serogroup O166 because of the similarity of properties with those of known colonization factors.

McConnell *et al.* (1990) showed that an ETEC strain of serotype O114:H21, which produced only heat-labile enterotoxin gave MRHA with bovine erythrocytes. One strain, E20738A, was shown to possess fimbriae of approximately 7.5 nm in diameter. Two possible fimbrial polypeptides, of molecular masses 17.5 and 15.5 kDa, were seen on SDS-PAGE. The 17.5 kDa band was the more prominent. The term CS17 has been used to describe this fimbria.

Aubel *et al.* (1991) identified a new adhesive factor, designated as Antigen 8786, which was isolated from the ETEC strain 8786 of serotype O117: H4. The coding region of antigen 8786 was located on a 70 MDa plasmid and has a molecular mass of 16.3 kDa. Antigen 8786 lacks Cys residues. The amino-terminal sequence showed no homology with those of other ETEC adhesive factors.

1.8 Aims of this thesis

In recent years, considerable effort has been made to identify and characterise fimbriae present on human ETEC strains, and to clone the genes involved in fimbrial biogenesis. Novel control mechanisms for expression of fimbriae have been revealed.

Much is still required to be done to reveal and understand the precise nature of the control mechanisms and the number of accessory proteins involved in the biogenesis of fimbriae on human ETEC.

The purpose of the work undertaken in this thesis was to expand the basic knowledge of the molecular organization and conduct a functional analysis of the various genes involved in the biogenesis of CS3 pili. This work is based on that reported by Manning *et al.* (1985), who have cloned and characterized plasmids encoding the DNA required for biogenesis of CS3 pili in *E. coli* K-12. The specific aims of this thesis are:

- To identify and sequence the minimal region encoding all genes necessary for CS3 biosynthesis in *E. coli* K-12.
- To identify the genes by sequence analysis, followed by experimental verification of the predictions.
- To analyze the transcription pattern of the genes contained within this region and identify the number and sizes of transcripts present in the CS3 operon.
- To investigate the possible roles of the various gene products and develop a model for CS3 biosynthesis.

CHAPTER 2

Materials And Methods

CHAPTER 2

Materials And Methods

2.1 Bacterial strains

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

2.2 Bacteriophage strains

M13mp18 and M13mp19 (Messing and Vieira, 1982) were used for cloning the DNA fragments for sequencing and oligomutagenesis.

2.3 Plasmids

Relevant features of the plasmid clones and vectors used in this study are listed in Table 2.2.

2.4 Media and Buffers

All media and buffers were prepared with Milli Q water and sterilised by autoclaving. Stock solutions of amino acids and antibiotics were added from sterile solutions after the media had been autoclaved and cooled to 45°C.

Composition of media and buffers that follow is given in grams per litre of various components.

2.4.1 CFA broth

Casamino acids	10 g
Bacto yeast extract	1.5 g
$MgSO_4.3H_2O$	0.05 g
$MnCl_2.4H_2O$	0.005 g
Water	to 1000 ml

2.4.2 L Broth

Sodium chloride	10 g
Bacto tryptone	10 g
Yeast extract	5 g
Water	to 1000 ml

The pH was adjusted to 7.0 before autoclaving.

Table 2.1

Bacterial Strains

Strain	Genotype/phenotype	Source/Ref.
C600	F$^-$ *thr-1 leuB6 tonA21 lacY1 supE44 thi-1 rfbD1*	Bachman (1987)
DS410	F$^-$ *minA minB sup E rspL*	Dougan and Sherratt (1977)
JM101	F' [*traD36 lacI^q* Δ*(lacZ)*M15 *proAB*]M15 *supE thi*-1 Δ*(lac-proAB)*	Yanish-Perron *et al.* (1985)
JM109	F' [*traD36 lacI^q* Δ*(lacZ)*M15 *proAB*] *recA1 endA1 gyrA96* (Nal^R) *thi hsdR17* (r_K-m_K+)*supE44 relA1* Δ*(lac-proAB)*	Yanish-Perron *et al.* (1985)
R594	W3350 F- *lac3350 galK2 galT22 rpsL179* Su$^-$	Bachman (1987)
TG1	F' [*traD36 lacI^q* Δ*(lacZ)*M15 *proAB*] *(lac pro) supE thi hsdD5*	Amersham
CB806	F- Δ*(lacZ) galK rps4 thi recA55 phoA8*	Schneider and Beck (1986)
SM10	RP4-2-*Tc::Mu thi thr leu supE*	Simon *et al.* (1983)
PC2251	*E. coli* C *minA minB supE*	J. B. Egan
PC2251	*E. coli* C *minA minB* Su$^-$	J. B. Egan
PB176	E13734/78 *E. coli* O6:H16 ST$^+$ LT$^+$ CFA/II$^+$ PB176	B. Rowe
PB176p$^-$	E13735/78 *E. coli* O6:H16 ST$^-$ LT$^-$ CFA/II$^-$	B. Rowe

Table 2.2

Vectors and Plasmid Clones

Plasmid	Characteristic	Reference
pBR322	Ap, Tc	Bolivar *et al.* (1977)
CDM8	*supF* gene	Seed (1987)
pMac 5-8	*bla* am	Stanssens (unpublished)
pCB182	Ap; promoter-detection vector	Schneider and Beck (1986).
pCB192	Ap; promoter-detection vector	Schneider and Beck (1986).
pSUP202	Ap; Cm; mobilizable vector	Simon *et al.*, 1983
pPM744	pSUP202+ *Hin*dIII fragment	This Thesis
pPM788	*Pst*I mutant	This Thesis
pPM 4024	UAG_{3523} mutant	This Thesis
pPM4025	*Sac*I mutant	This Thesis
pPM1351	*Nco*I mutant	This Thesis
pPM1352	*Bgl*II deletion mutant	This Thesis
pPM1353	*Eco*RI mutant	This Thesis
pPM 4017	*Apa*I deletion mutant	This Thesis
pPM791	UAG_{3578} mutant	This Thesis
pPM 4026	pMac 5-8 + *Hin*dIII fragment	This Thesis
pPM 4027	CDM8 + K^R cartridge	This Thesis
M13mp18	M13 vector	Messing and Vieira (1982)
M13mp19	M13 vector	Messing and Vieira (1982)

2.4.3 Nutrient Broth

Nutrient Broth (Difco)	16 g
Sodium Chloride	5 g
Water	to 1000 ml

2.4.4 2 x YT Broth

Sodium chloride	5 g
Bacto tyrptone(Difco)	16 g
Yeast extract (Difco)	10 g
Water	to 1000 ml.

The pH was adjusted to 7.0 before autoclaving.

2.4.5 M13 minimal medium

K_2HPO_4	10.5 g
KH_2PO_4	4.5 g
$(NH4)_2SO_4$	1.0 g
Sodium citrate	0.5 g
Water	to 1000 ml

Added the following after autoclaving and cooling to 45°C:

20% $MgSO_4$	1 ml (autoclaved)
20% glucose	10 ml (filter sterilised)
1% thiamine HCl	0.5 ml (filter sterilised)

2.4.6 TFB I (Transformation Broth I)

30mM KOAc

100mM RbCl

10mM $CaCl_2$

50mM $MnCl_2$

5% (v/v) Glycerol

The pH was adjusted to 5.8 with 0.2M acetic acid, and sterilised by filtration.

2.4.7 TFB II

10mM MOPS

75mM $CaCl_2$

10mM RbCl

15% Glycerol

The pH was adjusted to 6.5 with 1 M KOH, and sterilised by filtration.

2.4.7.1 Transformation component #a

Bacto yeast extract	5.0 g
Bacto Tryptone	20.0 g
Bacto Agar	14.0 g
$MgSO_4$	5.0g

Adjusted the pH to 7.6 with 5M KOH and sterilised by filtration.

2.4.7.2 Transformation component #b

Transformation components #b is same as transformation component #a without agar.

2.4.8 CFA agar plates

CFA broth + 2.0% bacto agar.

2.4.9 M13 minimal plates

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

2.4.10 Nutrient Agar plates

Lab-Lemco powder (Oxoid)	10 g
Peptone	10 g
Agar (Media Makers)	15 g
Sodium chloride	5 g
Water	to 1000 ml

NA plates containing antibiotics were prepared either by adding the appropriate antibiotics from sterile stock solutions to the medium before pouring the plates or by spreading the antibiotics on plates just before use. The final concentrations of the antibiotics were: ampicillin (Ap) - 50 μg/ml; chloramphenicol (Cm) - 30 μg/ml; tetracycline (Tc) - 16 μg/ml.

Incubations were at 37°C unless otherwise specified. Normally the liquid cultures were grown in 20 ml McCartney bottles.

2.4.11 McConkey agar plates

McConkey agar base (Difco)	40g
water	to 950ml

Filter-sterilised solutions of 20% galactose or lactose, as the case may be, was added after autoclaving and cooling to 45°C.

2.4.12 YT Soft agar

Yeast extract	5 g
Bacto Tyrptone	8 g
Sodium chloride	5 g
Bacto agar	7 g
Water	to 1000ml

The pH was adjusted to 7.0 before autoclaving.

2.5 Buffers

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

2.5.1 10 x TM

Tris-HCl pH 8.0	0.1M
$MgCl_2$	0.1M

2.5.2 10 x TE

Tris-HCl pH 8.0	0.1 M
EDTA	1.0 mM

2.5.3 10 x TAE

Contained 0.89 M Tris-acetate and 10 mM EDTA, pH 8.2, and was prepared by dissolving at room temperature,

Trizma base	48.2 g
Sodium acetate	16.4 g
EDTA	3.36 g
Water	to 1000 ml

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

2.5.4 10 x TBE

The 10 x TBE Contained 0.89 M Tris-borate and 10 mM EDTA, and was prepared by dissolving at room temperature,

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

The pH was adjusted to 8.3 before autoclaving.

2.5.5 TBS for Western blotting

Tris-HCl pH 7.4	20 mM
NaCl	0.9% (w/v)

2.5.6 TTBS for Western blotting

Tween 20	0.05% (v/v)
Tris-HCl pH 7.4	20 mM
NaCl	0.9% (w/v)

2.5.7 Transfer buffer for Western blotting

Trizma base	10.59 g
Glycine	50.45 g
Methanol	175 ml
Water	to 3.5 litres

2.5.8 10 X PBS

NaCl	80 g
KCl	2 g
Na_2HPO_4	11.5 g
KH_2PO_4	2 g
Water	to 1000 ml.

2.5.9 TMACl solution

for 200 mls

Tetra methyl ammonium chloride	65.76g
1 M Tris pH 8.0	10ml
0.1M EDTA	4ml
10% SDS	2ml
Water	to 200 ml

2.6 Chemicals

Acrylamide and mixed bed resin were from Bio-RAD. Bis (N,N'-methylene-bis-acrylamide), low melting point agarose and ultra pure urea were from Bethesda Research Laboratories (BRL). Sea plaque (Seakem) agarose for routine agarose gel electrophoresis was purchased from Sigma. Bacto yeast extract, tryptone, McConkey agar base and bacto agar were from Difco. Agar from Media Makers, South Australia, was used in the routine preparation of L plates. Adenosine-5'-triphosphate (ATP), herring sperm DNA, dithiothreitol (DTT), sodium dodecyl sulphate (SDS), N,N,N',N'-tetramethylenediamine (TEMED), ampicillin and chloramphenicol were from Sigma. Ammonium persulphate (APS) was from May and Baker.

BCIG (5-bromo-4-chloro-3-indolyl-3-β-d-Galactopyranoside), iso-propyl β-D-thiogalactopyranoside (IPTG), Trizma base, deoxy-ribonucleotide triphosphates (dATP, dCTP, dGTP and dTTP) and the dideoxy-ribonucleotide triphosphates (ddATP, ddCTP, ddGTP and ddTTP) were from Boehringer-Mannheim. Molecular biology grade

caesium chloride was from Bethesda Research Laboratories (BRL), whereas the technical grade caesium chloride was from BDH.

Ethylene-diamine-tetra-acetic-acid (EDTA) was from Ajax Chemicals.

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

Polyethylene glycol-6000 (PEG), also purchased from BDH, was used for preparing M13 single stranded phage for sequencing.

The sequencing primer (17mer; 5'-GTAAAACGACGGCCAGT-3') was obtained from BRESATEC (Adelaide). The -35 sequencing primer was obtained from New England Biolabs.

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

2.7 Enzymes

Restriction endonucleases were routinely purchased from New England Biolabs, Boehringer-Mannheim, Pharmacia or Amersham.

DNase I, RNAse-A and lysozyme were obtained from Sigma. Stock solutions of RNAse-A (10 mg/ml) were heated to 80°C to inactivate any contaminating DNAses.

Bovine Serum Albumin (BSA, enzyme grade) was obtained from BRESATEC (Adelaide). Solutions of BSA at 2mg/ml were prepared in water and stored at -20°C .

Other enzymes were purchased from the following suppliers: New England Biolabs (T4 DNA ligase), Amersham (T4 DNA polymerase), Boehringer-Mannheim (Klenow and Calf intestinal alkaline phosphatase) and IUB (Sequenase TS).

2.8 Radionucleotides

α-[^{32}P]-dCTP at a specific activity of 1,700 Ci/mMole and γ-[^{32}P]-dATP (2399Ci/mmol) were obtained from BRESATEC (Adelaide). [^{35}S]-Methionine (1,270 Ci/mMole) and [^{35}S]-dATP (>1000 Ci/mMole) were purchased from Amersham. [^{14}C]-galactose # CFA 435 was purchased from Amersham.

2.9 Miscellaneous materials

Kodak X-ray film was used for autoradiography. Polaroid film No. 667 from Polaroid Film Co. was used for taking photographs of agarose gels. Dialysis tubing was from Union Carbide. Ordinary filter papers and GF/A filters were from Whatmann. Micro porous filters were from Millipore.

2.10 Synthesis of oligodeoxynucleotides

Oligo-deoxynucleotides were synthesised using reagents purchased from Applied Biosystems or Ajax Chemicals (acetonitrile). Synthesis was performed on an Applied Biosystems 381A DNA synthesizer.

2.11 General techniques

2.11.1 Maintenance of bacterial strains

All strains were maintained as lyophilised cultures (see below) and stored *in vacuo* in sealed glass ampoules. When required, an ampoule was opened and its contents resuspended in several drops of the appropriate sterile broth and used in starting liquid or plate cultures. Single colonies from the plates were selected for subsequent use or storage. Short term storage of strains in routine use was as a suspension of freshly grown bacteria in 32% (v/v) glycerol, containing 0.6% (w/v) peptone, at -70°C, or in 40% glycerol, containing 0.6% (w/v) peptone, at -20°C.

2.11.2 Lyophilization of bacterial cultures

Several colonies were suspended in a small volume of sterile skimmed milk. Approximately 0.2 ml each of this bacterial suspension was dispensed into sterile 0.25 X 4 inch freeze-drying ampoules and the end of each ampoule was plugged with cotton wool. The samples were then lyophilised in a freeze drier. After releasing the vacuum the cotton wool plugs were pushed well down the ampoule and a constriction was made just above the level of the plug. The ampoules were evacuated to a partial pressure of 30 microns, sealed at the constriction without releasing the vacuum and stored at 4°C.

2.11.3 Purification of colonies and plaques

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

2.12 Analysis and manipulation of DNA

2.12.1 Small scale plasmid DNA extraction

Plasmid DNA was isolated by one of the following procedures.

2.12.1.a Method 1

Small scale plasmid purification from 1 to 10 mls of cultures were done essentially by the three-step alkali lysis method of Birnboim and Doly (1979), as described by Maniatis *et al.* (1982), with some minor modifications. Overnight bacterial cultures (1.5 ml) were transferred to a microfuge tube, harvested by centrifugation (10 minutes, Eppendorf) and resuspended in 0.1 ml of GTE solution (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA). The cells were lysed by adding 0.2 ml of a freshly prepared solution of 0.2 M NaOH, 1% (w/v) SDS followed by a 5 minute incubation at room temperature. Proteins and chromosomal DNA were precipitated by the addition of 0.15 ml of KAc solution (5M potassium acetate, pH 4.8). After a 5 minute incubation at room temperature, and centrifugation in an Eppendorf for 5 minutes, the supernatant (~400 µl) was transferred to a fresh tube and precipitated by the addition of 1ml of ethanol. The DNA was collected by centrifugation, washed with 95% (v/v) ethanol and dried *in vacuo*. The pellet was resuspended in 200 µl of TE+RNAse (5µg/ml) and, after a 5 minute incubation at 37°C, the DNA solution was extracted with TE-equilibrated phenol. The DNA was further ethanol precipitated and resuspended in 20 µl of TE. Usually 2-3µl of the DNA solution was used for restriction analysis.

2.12.1.b Method 2

Small scale plasmid purification from 1.5 ml overnight cultures was by a modification of the alkali lysis method, as suggested by Garger *et al.* (1983), wherein the supernatant after KAc precipitation step in Method I was transferred to a fresh tube and

extracted once with TE-equilibrated phenol. Plasmid DNA was precipitated by the addition of 2 volume of 95% ethanol followed by a 2 minute incubation at room temperature and centrifugation. The DNA pellet was washed with 70% (v/v) ethanol, dried *in vacuo* and resuspended in 40 µl of TE.

2.12.1.c Method 3

Small scale plasmid DNA was also prepared by the Triton X-100 (cleared lysate) method of Kahn *et al.* (1979): Cells from 10ml overnight cultures were resuspended in 0.4 ml of 25% (w/v) sucrose in 50 mM Tris-HCl pH 8.0. After adding 50µl of lysozyme (10 mg/ml; freshly prepared in H_2O) and 50 µl of 0.25M EDTA, pH 8.0, the cells were left on ice for 15 minutes. Then 0.5 ml of the TET buffer (50 mM Tris-HCl, 66 mM EDTA, pH 8.0, 0.4% Triton X-100) was added and mixed by inverting the tubes a few times. The chromosomal DNA was pelleted by centrifugation (20 minutes, 4°C, Eppendorf) and the supernatant was extracted with TE-saturated phenol, followed by one extraction with diethyl-ether. Plasmid DNA was precipitated by the addition of an equal volume of propan-2-ol and cooled at -70°C for 30 minutes. The precipitate was collected, washed once with 1 ml of 70% (v/v) ethanol, dried in a vacuum desiccator for 10 minutes. and resuspended in 50 µl of TE.

2.13 Large scale plasmid DNA isolation

2.13.1.a Method 1

Large scale plasmid DNA preparations were performed according to the method of Birnboim and Doly (1979) as described by Maniatis *et al.* (1982) with the exception that the nucleic acids were precipitated with ethanol. The pellet was then resuspended in 8 mls of TE and 8 grams of caesium chloride was dissolved in it. After adding 200µl of an ethidium bromide solution (10mg/ml in TE) the DNA/CsCl solution was centrifuged to equilibrium in a Beckman Ti50 rotor at 45,000 r.p.m. for 42 hours at 15°C. The DNA bands were visualised in ordinary light and the lower band was collected. The ethidium bromide was removed by extracting three times with iso-propanol, which had been

equilibrated with 5M NaCl, 10mM Tris-HCl pH8.0 and 1mM EDTA, and dialysed at 4°C against one litre of TE. The buffer was changed 3 to 4 times at intervals of 4 to 12 hours.

2.13.1.b Method 2

Large scale plasmid purification was also performed by the three step alkali lysis method of Garger *et al.* (1983). Cells from a one litre overnight culture were harvested and resuspended in 24 ml of solution-I (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). Freshly prepared lysozyme (4 ml of 20 mg/ml in solution-I) was mixed with the cell suspension and incubated at room temperature for 10 minutes. Addition of 55 ml of solution-II (0.2 M NaOH, 1% (w/v) SDS) followed by a 5 minute incubation on ice resulted in lysis of the cells. To this was added 28 ml of solution-III (5M potassium acetate, pH 4.8) and incubated on ice for 15 minutes. The protein, chromosomal DNA and high molecular weight RNA were removed by centrifugation at 8,000 rpm at 4°C for 20 minutes. The supernatant was then extracted with an equal volume of TE-saturated phenol: chloroform: iso-amyl alcohol mixture in the ratio of 25:24:1. Plasmid DNA from the aqueous phase was precipitated with 0.6 volume of propan-2-ol at room temperature for 10 minutes and collected by centrifugation at 10,000 rpm at 4°C for 35 minutes. After washing with 70% (v/v) ethanol the pellet was dried in a vacuum desiccator for 15-20 minutes. The DNA was resuspended in 4.8 ml of TE and purified by centrifugation on a two-step (1.8 and 1.47 g/ml) CsCl/ethidium bromide gradient. The DNA band was recovered and the ethidium bromide removed by extracting with iso-amyl alcohol. CsCl was then removed by dialysis against TE at 4°C in 5 litres of TE overnight. The buffer was changed 3 to 4 times at intervals of 4 -12 hours and the DNA was stored at 4°C.

2.14 Phenol extraction of nucleic acids

2.14.a Equilibration of phenol

Molecular biology grade phenol from BDH was equilibrated as follows: an equal volume of 1M Tris pH8.0 and 5mg of 5-hydroxy-quinoline were added to the melted phenol and heated in a 65°C oven until the aqueous and organic phases became one

(usually took 1 to 2 hours). This was followed by cooling to room temperature in the dark to separate the two phases. Aliquots of 10 mls of phenol phase from this master mix, stored at -20°C, were mixed with 10 ml of TE and let stand at room temperature until the phases separated. The TE phase was then removed and the whole procedure repeated twice. This working stock of phenol was stored under TE at 4°C. Fresh working stock was prepared at every 3 to 4 week intervals.

2.14.b Extraction with phenol

Extraction with phenol was done by mixing DNA-containing solutions with 0.5 volume of equilibrated phenol, cooling on ice and centrifuging at room temperature or at 4°C. The aqueous phase was withdrawn and the procedure was repeated, if necessary, until no protein band was visible at the interphase. Finally, the aqueous phase was ethanol precipitated and the DNA resuspended in TE.

2.15 Ethanol precipitation

Plasmids and linear double stranded DNAs were precipitated by adding NaCl to 200mM and 2.5 volume of ethanol, followed by freezing either at -70°C for 30 minutes or in a dry ice/ethanol bath for 15 minutes. For the recovery of small quantities of DNA, 2 to 5µl of glycogen (20mg/ml; Boehringer-Mannheim) was added as carrier. The precipitated DNA was recovered by centrifugation in Eppendorf centrifuge for 15 minutes at room temperature or at 10K (10,000 rpm) for 30 minutes at 4°C (for large volumes precipitated in 50ml Oak Ridge tubes). The pellet was then washed once with 95% ethanol, dried in a vacuum desiccator for 10 minutes and dissolved in TE.

Single stranded M13 DNA was prepared by phenol extracting the phage suspension in TE and precipitating the aqueous phase with 2.5 vol. of ethanol after adding 300 mM sodium acetate pH 4.8, instead of the 200mM NaCl used in precipitating double stranded DNA.

2.16 DNA quantitation

Concentrations of DNA in solutions were determined either by measuring the absorption at 260 nm, assuming that an A_{260} of 1.0 represents 50 µg/ml of DNA (Miller, 1972), or by electrophoresing on agarose minigel and comparing the intensities of ethidium bromide-stained bands with the intensities of bands containing known concentrations of DNA standards.

2.17 Restriction Endonuclease Digestion

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

DNA+ water	15µl (up to 10µg of DNA)
10 x digestion buffer*	2µl
BSA (1mg/ml)	1µl
Restriction enzyme	2µl (0.6 to 20 units)

* Separate buffers for each enzymes was made according to the assay conditions described by the manufacturers' recommendations and stored at -20°C.

Reactions were terminated by adding EDTA to final concentration of 30mM and heating to 65°C for 10 minutes.

Those enzymes which required similar concentration of NaCl were used simultaneously when needed to digest with more than one enzyme. When the optimal salt concentrations of enzymes differed the one which required the lowest NaCl concentration was used first and, after an hour's incubation, were added the other(s) after supplementing the mix with the required amount of NaCl. When the reaction buffers differed considerably, the DNA was ethanol precipitated after the first digestion and resuspended in the appropriate digestion buffer before proceeding with the second digestion.

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

2.18 Size determination of DNA fragments

The sizes of restriction fragments were calculated by comparing their relative mobility with the molecular weight markers, such as the *Eco*RI digested DNA of the *Bacillus subtilis* bacteriophage SPP1. The sizes of these SPP1 *Eco*RI fragments differed from those published (Ratcliff *et al.*, 1979) and were initially re-calculated using *Hind*III digested bacteriophage lambda DNA and *Hpa*II digested plasmid pUC19 DNA as standards. The newly estimated sizes (kb) are: 8.51; 7.35; 6.11; 4.84; 3.59; 2.81; 1.95; 1.86; 1.51; 1.39; 1.16; 0.98; 0.72; 0.48; 0.36 (BRESATEC, Adelaide). The plasmid pEC701 (Sivaprasad, *et al.*, 1990), digested with *Hin*FI, was used as small size standard (sizes in bp: 1464, 1378, 783, 690, 517, 509, 506, 489, 396, 344, 319, 314, 298, 221, 220, 154, 132, 119, 94, 81, 75, 74 and 32).

2.19 Analytical and preparative separation of DNA fragments

Electrophoresis of digested DNA was carried out at room temperature on 0.6%, 0.8% or 1% (w/v) agarose gels, 13 cm long, 13 cm wide and 0.7 cm thick in a horizontal gel apparatus. Gels were run at 100V for 4-5 hr in either TBE or TAE buffer. The gels were stained with ethidium bromide (0.0004% w/v in TBE or TAE) and photographed under short wave UV, using either Polaroid 665 or 667 film.

For preparative gels low melting point agarose at a concentration of 1% (w/v) was used and the DNA bands recovered by the following methods:

2.19.a Method 1

The gel slices containing the required DNA bands were excised into Eppendorf tubes and approximately 2 volumes of 50mM Tris-HCl pH 8.0 + 0.5mM EDTA were added. The gel was melted by at 65°C for 10 minutes. After cooling to 37°C the DNA was recovered by two phenol extractions followed by ethanol precipitation, by adding 2µl of glycogen (20 mg/ml) as carrier and 300mM of sodium acetate pH 5.0.

2.19.b Method 2

Gel slices containing the required DNA bands were excised and then placed inside a dialysis bag. About 400µl of TE was added, ends tied such that the bag was in a fully

inflated state, placed across the path of electric current in a horizontal gel apparatus and electrophoresed at 100mA for 5 minutes (10-20 minutes when the length of DNA fragment was more than 4 kb). The bag was inverted a few times to dislodge any DNA that might be sticking to the walls, and the buffer was taken out by piercing the side of the bag and the DNA was precipitated with ethanol.

2.19.c Method 3

Gel slices containing the required DNA bands were excised and placed in an Eppendorf tube, containing siliconized glass wool covering a hole at the bottom of the tube. This was inserted into another Eppendorf tube and centrifuged for 15 minutes at 6500 r.p.m. The DNA contained in the elute was collected by ethanol precipitation as previously described.

2.20 Polyacrylamide gel electrophoresis

2.20.1.a Non-denaturing PAGE

The gel stock (30%) was prepared by dissolving recrystallised acrylamide and bis in a 50:1 ratio in distilled water at room temperature. This was de-ionised by stirring with 5g/100ml of mixed-bed resin (Bio-Rad) for 30 minutes at room temperature. After removing the resin by filtration through scintered glass funnel the gel stock was de-gassed by applying vacuum for 30 minutes and stored at 4°C in the dark.

Gels were poured and used on the same day. A 20 x 40cm gel was prepared by mixing 10mls of 30% stock solution, 6mls of 10 x TBE, 385µl of 25% APS and 95µl of TEMED and poured into the gel sandwich, which had been pre-warmed to 37°C. The gel was allowed to polymerise at 37°C for at least 30 minutes before use. Immediately after removing the comb, the wells were rinsed with distilled water and the gel was pre-electrophoresed at 100V for 15 minutes before loading the DNA. Electrophoresis was done at a constant 400 to 500V. The bands of DNA fragments, end-labelled before electrophoresis, were visualised by autoradiography at room temperature.

2.20.1.b Denaturing (sequencing) gel

The sequencing gel stock (6% polyacrylamide, 8M Urea in TBE) was prepared by dissolving, at 37°C, 57g acrylamide, 3g bis-acrylamide and 480g urea in 400 ml of distilled water. This was de-ionised by stirring 35g mixed-bed resin for 30 to 60 minutes and filtered through scintered glass funnel. After adding 100 mls of 10xTBE the volume was made upto 1 litre with water and degassed under vacuum for 2 hours. The gel stock was stored at 4°C in the dark and used for a maximum period of 2-3 months.

For pouring a 28 x 40 x 0.025 cm gel, 75ml of the gel stock was mixed with 480µl of freshly made 25% ammonium persulphate solution and 120µl of TEMED. After allowing to polymerise for 30 to 60 minutes the gel was set up the apparatus, and a metal plate was clamped to the front of the gel to maintain a uniform temperature across the gel surface. The gel was pre-electrophoresed for 30 minutes at 1000V, using TBE as the running buffer. The comb was left in place while pre-electrophoresis, as prolonged electrophoresis without the comb resulted in distortion of the wells. The comb was removed, and the wells flushed with buffer to remove the unpolymerized acrylamide, just before loading the reaction mixes.

2.21 Elution of DNA from polyacrylamide gel

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

2.22 Dialysis

Dialysis bags were prepared by boiling the tubing (18/32) in TE for 5 minutes, followed by washing the inside of the bags with the same buffer at room temperature. Dialysis tubes were prepared and used on the same day. Dialysis was done at 4°C in

1 litre of TE with 3 to 4 changes of the buffer at intervals of 4 to 12 hours. The DNA solution was then ethanol precipitated and resuspended in TE.

2.23 De-phosphorylation of DNA

Restriction enzyme digested DNA was de phosphorylated by either of the following two methods:

2.23.a Method 1

Approximately 0.5-1.0 µg of digested plasmid DNA was incubated, in a 60 µl volume containing 100 mM Tris-HCl pH 8.0, 0.17% SDS, with 1-5 units of calf intestinal alkaline phosphatase (CIP; Boehringer-Mannheim) for 20 minutes at 37°C. The reaction was terminated by the addition of EDTA, pH 8.0 to a final concentration of 30 mM followed by heating at 70°C for 10 minutes. The DNA was extracted twice with TE-saturated phenol and precipitated overnight at -20°C with two volumes of ethanol and 300mM of sodium acetate pH 5.0.

2.23.b Method 2

Approximately 0.5-1.0 µg of digested plasmid DNA was placed at 65°C for 10 minutes to heat inactivate the enzyme. The 100µl digested DNA aliquot was made to a final volume of 290µl with 30µl of 1M Tris-HCl pH 8.0, 5ml of 10% SDS, 5µl of CIP (1u/µl) and 150 µl of water. After incubating at 37°C for 2 hrs. the CIP was inactivated by incubating at 70°C for 10 minutes and removed by phenol extraction. The DNA was recovered by ethanol precipitation.

2.24 End-labelling and end-filling

End-labelling, used in identifying small DNA fragments (less than 1kb) on polyacrylamide gels, was done as follows:

α–[^{32}P]-dATP (or α–[^{32}P]-dCTP)	4µl (vacuum dried)
DNA + water	8µl
10 x TM	1µl
Klenow enzyme	1µl (1unit)

After incubating the above mix at 37°C for 15 minutes, 2µl of dNTP solution (0.25mM each of dATP, dCTP, dGTP and dTTP in 5mM Tris-HCl pH8.0, 1.0mM EDTA) was added and the incubation continued for a further 15 minutes. The reaction was terminated by adding EDTA to 30 mM or heating to 70°C for 10 minutes.

End-filling was essentially similar to end-labelling except that the radionucleotide was omitted. Addition of EDTA, as well as heat-killing of Klenow, was omitted when the DNA was to be used for ligations.

2.25 Ligation

A typical ligation reaction contained 20 to 100 ng of DNA, 5mM rATP, 10mM $MgCl_2$, 10mM dithiothreitol, 50mM Tris-HCl pH 7.5 and 0.2 units (for sticky end ligation) to 2 units (for blunt end ligations) of T4 DNA ligase in a 20µl reaction volume. Polyethylene glycol (PEG 6000; BDH) was added to 25% to increase the efficiency of ligations when the DNA concentrations were low. Ligations were done at 14°C for 1 to 16 hours.

For M13 cloning, a vector to insert molar ratio of 1:3 was used. For this, the amount of vector was kept constant, at 20ng per ligation, and the insert DNA was added at approximately 10 ng/kb of fragment's length.

2.26 Transformation of *E. coli*

2.26.a $CaCl_2$ Method

Transformation was performed essentially according to the method of Brown, *et al.* (1979). *E. coli* K-12 strains were made competent for transformation with DNA as follows:

An overnight culture in LB (or M13 minimal medium for JM101/TG-1) was diluted 1:100 into LB (or 2xYT for JM101/TG-1) and incubated with shaking at 37°C until an A_{600} of 0.6. The cells were chilled on ice for 20 minutes, sedimented (5000 r.p.m. at 4°C for 5 minutes.) and resuspended in 0.5 vol of cold 100 mM $MgCl_2$. The cells were pelleted again and resuspended in 1 to 2 ml of ice-cold 100 mM $CaCl_2$. This

was allowed to stand for at least 1 hour on ice before adding the DNA. Competent cells were prepared and used on the same day.

Transformation was done by mixing 5 to 20 ng of DNA and 0.2 ml of competent cells in a chilled sterile screw-capped polycarbonate tube. After keeping on ice for 30 minutes, the cells were heat-shocked for 5 minutes at 37°C and chilled a further 10 minutes on ice. The cells were then diluted by adding 3 ml of LB and incubated with shaking at 37°C for 1 hr. Fractions of 0.1 ml were spread directly on selective plates, or the entire culture was concentrated by centrifugation and plated. Competent cells not transformed with DNA were included as control.

2.26.b RbCl Method (M. Scott and V. Simanis personal communication)

E. coli K-12 strains were made competent as follows: the bacterial stock culture was streaked out onto a transformation component #a plate (section 2.4.7.1) and incubated overnight at 37°C. A single colony was inoculated into 5 ml of transformation component #b (section 2.4.7.2) medium and incubated to an A_{650} of 0.6. This was diluted 1:20 into 100 ml of transformation component #b medium and incubated with shaking for a further 2 hours until it reached an A_{650} of 0.5 - 0.6. (4x 10^8 cells/ml). The cells were chilled on ice for 5 minutes, pelleted at 4°C and resuspended in 2/5 vol. of TFB-I (section 2.4.6). This was allowed to stand for 5 minutes on ice, cells pelleted again for 5 minutes at 6000 r.p.m. at 4°C and resuspended in 1/25 vol. of TFB-II (section 2.4.7). After standing on ice for a further 15 minutes, aliquots of 200 µl were snap-frozen and stored at -70°C.

For transformation, competent cells stored at -70°C were allowed to thaw on ice and 0.1 ml was mixed with DNA and left on ice for 45 minutes. The cell/DNA mixture was heat shocked at 42°C for 90 seconds and kept on ice for 2 minutes, before adding 2 ml of LB and incubating with shaking at 37°C for 45 minutes. An aliquot of 0.1ml was then plated onto selection plates directly.

2.27 M13 transfection

Transfection was performed by mixing 1µl of ligated DNA with 0.2ml of competent cells in chilled sterile glass tubes and keeping on ice for 40 minutes. After heat shock treatment at 45°C for 2 minutes, the cells were plated on M13 minimal plates with 3ml of molten YT soft agar, containing 20µl of IPTG (24mg/ml stock solution made in water), 20µl of BCIG (20mg/ml stock made in dimethyl formamide) and 0.1ml of exponential phase cells. Incubation was at 37°C overnight.

2.28 Mobilization of plasmid into ETEC cells

The plasmid to be mobilized was transformed into the donor strain, SM10 (Simon *et al.*, 1983; Table 2.1), and selected on chloramphenicol plates. The donor and recipient strains, SM10 and PB176p$^-$ (CS3$^-$), respectively, were grown overnight in NB or NB, containing chloramphenicol. After diluting 1:20 in the same medium, and growing to early exponential phase with slow agitation (not more than 200 r.p.m), the donor and recipient bacteria were mixed at a ratio of 1:10 and allowed to stand at 37°C for 3-4 hours. The mixture was plated onto minimal plates containing chloramphenicol and incubated overnight at 37°C.

2.29 Labelling of the oligonucleotides

Oligonucleotides were kinased using a reaction mix consisting of 1µl of primer (~60-100 ng), 1µl 100 mM DTT, 1 µl of 10 x kinase buffer (500 mM Tris-HCl pH7.4 and 100 mM $MgCl_2$), 1µl of T4 polynucleotide kinase (2-3 units/µl) and 10 µl of lyophilised γ-[^{32}P]-dATP in a total volume of 10µl. After a 30 minute incubation at 37°C the labelled oligonucleotide was purified either by PAGE or by removal of the unincorporated dNTPs by centrifugation through a Sepharose CL-6B column. The oligonucleotide was heated to 65°C for 10 minutes before use, to destabilise the secondary structures.

Sepharose CL-6B columns were prepared by placing glass wool into an Eppendorf tube with a hole punched at the bottom. After adding 1 ml of Sepharose CL-6B (equilibrated with 1x TE buffer) the tube was placed within another tube and

centrifuged at 2000 r.p.m for 2 minutes in a bench centrifuge (Hermle). The column was washed twice with sterile distilled water before use.

2.30 Galactokinase Assay

The GalK assay was performed essentially as described by McKenney *et al.* (1981): an overnight culture of CB806 (Schneider and Beck, 1986; Table 2.1) harbouring the appropriate plasmid was subcultured in M9 minimal medium using fructose as a carbon source and grown at 37°C until the OD_{650} reached 0.6. A 1 ml sample was removed to a test tube to which 40µl of lysis buffer (100 mM EDTA, 100 mM DTT, 50 mM Tris-HCl pH 8.0) and a drop of toluene were added. After vortexing for 10 seconds the tubes were incubated at 37°C for 1hr. From this lysis mix a 20µl sample was removed to a clean test tube to which was added, 20µl of solution-I (5 mM DTT, 16 mM NaF), 50µl of solution-II (8 mM $MgCl_2$, 200 mM Tris-HCl pH 7.9, 3.2 mM ATP) and 10µl of ^{14}C-galactose (diluted to 4.5 x 10^6 dpm/mmole). This mixture was incubated at 32°C during which 25µl samples were removed at 0, 5, 10, 20 and 30 minutes intervals and spotted on 23mm DE81 Whatmann filters and air dried. Two samples were also spotted onto separate discs and retained unwashed. The other discs were batch washed 3 times at 5 minutes intervals in 300mls of water, dried at 65°C for 10 minutes and counted with 3 mls of Optiphase "Hisafe II" (LKB) scintillation fluid.

The specific activity of galactokinase (GalK) was determined according to the formula of McKenney *et al.* (1981).

Results are presented as mean values of three independent experiments in which GalK was assayed in duplicate .

2.31 Analysis and manipulation of RNA

2.31.1 Preparation of cellular RNA

2.31.1.a Method 1

RNA was prepared by a modified method of Chomczynski and Sacchi (1987) as follows. Overnight cultures were subcultured and grown to an A_{650} of 0.8 to 1.0. The

cells from a 50 ml culture were harvested and washed twice in 10 ml each of 1mM EDTA-PBS. To this was added 800µl of denaturing solution (solution D; 4M guanidinium thiocyanate, 2mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1 M β-mercaptoethanol), swirled the cell suspension gently, and sequentially added 0.1 ml of 2M sodium acetate pH 4.0, 1 ml of water-saturated phenol and 0.2 ml of chloroform. Thoroughly mixed the solution by inverting the tube after the addition of each reagent. After a 10 minutes' incubation at 4°C, followed by a 15 minute centrifugation, the aqueous phase was transferred to a fresh tube, mixed with an equal volume of iso-propanol and placed at -20°C for 1 hour to overnight. The precipitated RNA was sedimented at 10,000g for 20 minutes and dissolved in 0.5 ml of solution D. This was transferred into a 1.5 ml Eppendorf tube and precipitated again with 1 vol of iso-propanol at -20°C for at least 1hour. The pellet was resuspended in 75% ethanol, sedimented, vacuum dried and dissolved in 100µl of RNase-free water. If the RNA was not to be used immediately, it was mixed with 1/10 vol of 3M sodium acetate pH5.6 and 2.5 vol of ethanol and stored at -20°C.

2.31.1.b Method 2

RNA was also prepared by a modified method described by Aiba *et al.* (1981) as follows: overnight cultures were subcultured and grown to an A_{650} of 1 in NB. The cell pellet from 5 mls was resuspended in 0.5 ml of solution A (0.02 mM NaAc pH 5.5, 0.5% SDS, 1 mM EDTA) and extracted 3 to 4 times with hot (65°C) phenol which had been equilibrated with a solution containing 0.02 mM NaAc, 0.02mM KCl and 0.01mM $MgCl_2$ pH 5.2. After precipitating with two volumes of ethanol and 1/10 volume of NaAc, the contaminating DNA was removed by incubating at 37°C for 10 to 15 minutes with 1 ml of DNAse I (10u/µl, BRESATEC) in 20 mM Tris-HCl pH 7.6, 5mM $MgCl_2$. The RNA was re-extracted with phenol, ethanol precipitated three times, dried *in vacuo* and resuspended in water for immediate use or stored in ethanol for later use.

2.31.2 Quantitation of RNA

The concentration of RNA in solutions was determined by measuring the absorbance at 260nm, assuming an A_{260} of 1.0 is equal to 40 µg RNA/ml (Miller, 1972).

2.31.3 Preparation of single stranded probe

Strand-specific probes were made from M13 clones, containing the required DNA fragment, by using the hybridisation probe primer. The primer (2-3 ng) was annealed to the single strand template DNA (1µg) and extended with Klenow in the presence of all four dNTPs (0.25 mM each) to generate a partially double-stranded DNA, leaving the insert DNA in single stranded form. The probe synthesised this way was used without denaturing.

2.31.4 Hexamer labelling method to make double strand probe

Probes were also made from double stranded DNA using the random hexamer labelling method (BRESATEC, Adelaide). The reaction mix contained,

Template DNA (6-10ng/µl)	11µl
Hexamer nucleotide solution	6µl
dNTP mix + 10 x TM	6µl
α [^{32}P]-dCTP (10mCi/ml)	5µl

The above components were mixed together and heated to 100°C for 5 minutes and cooled to room temperature for 15-30 minutes. Klenow was added to the reaction mix and incubated at 37°C for 30 minutes. The probe prepared in this fashion was boiled for 3 minutes and snap-chilled in a dry ice/ethanol bath before adding to the pre-hybridisation solution.

2.31.4 Preparation of RNA Dot Blots

RNA solutions (1-5 µg in 100 µl H_2O) were denatured by the addition of 300µl of 6.15M formaldehyde/10xSSC and incubated at 65°C for 10 minutes. A nitrocellulose filter (Schleicher and Schüll) equilibrated with 10 x SSC was placed in the dot blot apparatus (BIO-RAD) and the denatured RNA was spotted. The filter was then air dried,

baked *in vacuo* at 80°C for 2 hr and subjected to hybridisation (section 2.31.5.d and 2.31.5.e).

2.31.5 Northern blot

Glyoxalation of RNA, electrophoresis and Northern transfer procedures (Thomas, 1980) were as described below:

2.31.5.a Glyoxalation of RNA

The RNA samples were denatured by incubation of RNA (20-30 µg) with 1M deionized glyoxal and 10mM sodium phosphate buffer pH 6.5 at 50°C for 60 minutes. The reaction mix was chilled on ice and 5µl of sample buffer containing 50% (v/v) glycerol, 10mM sodium phosphate buffer pH 6.5 and bromophenol blue (0.4%) was added.

2.31.5.b Electrophoresis

Denaturing electrophoresis of RNA was performed on 1.8% agarose gels in 10 mM phosphate buffer. Upto 20µg of RNA was loaded in each lane, after heating to 100°C for 2-3 minutes. Gels were run at 20V with circulating buffer, in order to keep the pH at acceptable limits (glyoxal dissociates from RNA at pH > 8.0), for the glyoxal adduct to remain associated with the RNA during electrophoresis .

2.31.5.c Transfer to the nitrocellulose filter

No treatment of the gel was needed before transfer. Transfers were set up in 20 x SSC (3M NaCl, 0.3M trisodium citrate). The RNA was transferred from gel to the nitrocellulose membranes by capillary blotting, as described by Maniatis *et al.* (1982). After 17 hours of transfer, the filters were air dried and baked at 80°C for 2 hours. Prior to pre-hybridisation, the glyoxal adduct was removed by treating with 20mM Tris pH 8.0 at 100°C and allowed to cool to room temperature.

2.31.5.d Pre-hybridization

Filters were pre-hybridized for 8-12 hours at 42°C in 50% (v/v) formamide, 50mM sodium phosphate buffer pH 6.5, 5 x SSC (0.34M NaCl, 75mM sodium citrate pH

7.0), 5 x Denhardt's solution, (for 50x Denhardt's: Ficoll 5g, Polyvinylpyrrolidone 5g, BSA fraction V 5g in 500ml water) and 250 µg /ml Herring sperm DNA (Note: The herring sperm DNA was boiled for 3-5 minutes at 100°C just before adding to the pre-hybridization mixture).

2.31.5.e Hybridization, washing of filters and autoradiography

Denatured probe was added to the pre-hybridized filter and incubated at 42°C for 20-48 hours. Filters were washed twice for 5 minutes at room temperature in 2 x SSC and 0.1% SDS (w/v). This was followed by two washes for 15 minutes each at 50°C in 0.1 x SSC and 0.1% SDS (w/v). After air drying for 15-20 minutes, the filters were covered in plastic wrap and autoradiographed at room temperature.

2.31.6 Primer extension

RNA (10µg) and the primers, 5' labelled by kinasing, were co-precipitated with ethanol and resuspended in 50µl of 10mM Tris-HCl (pH 8.0), 1mM EDTA and 1M NaCl. This solution was heated to 75-80°C for 3 minutes, to denature the RNA, and incubated at 42°C for 2hours to allow hybridisation to occur. After slow cooling to room temperature, the primed RNA was precipitated by the addition of 100µl of ethanol, the pellet dried and resuspended in 24µl of 50mM Tris-HCl (pH 8.3 at 42°C), 40mM KCl, 6mM $MgCl_2$, 1mM dithiothreitol and 500mM each of all four dNTPs. Reverse transcriptase (from Avian myeloblastosis virus; 20 units) was added and incubated at 42°C for 1hour. The reaction was stopped by extraction with phenol and precipitated with ethanol. The pellet was dried and re-dissolved in TE, containing 10mg/ml RNase A, and incubated for 30 minutes at room temperature. After a further extraction with phenol and two ethanol precipitations the pellet was dissolved in 5µl of TE. Formamide/dye loading buffer (5µl) was added and the sample was electrophoresed immediately after denaturation at 100°C for 3 minutes. A DNA sequencing ladder, generated by using the above primers, was run on the same gel as a size marker.

2.32 Protein analysis

2.32.1 Minicell labelling

Minicells from the *E. coli* strain DS410 (Dougan and Sherratt, 1977), carrying the appropriate plasmid, were purified and the plasmid-encoded proteins labelled with [^{35}S]-Methionine by the method of Kennedy *et al.* (1977) as modified by Achtman *et al.* (1979). The cells (250ml overnight culture in LB) were initially collected by centrifugation (Sorvall rotor GSA, 7,000 rpm, 10 minutes, at 0°C). Minicell purification involved separation of minicells from whole cells by centrifugation (Sorvall rotor HB4, 6,000 rpm, 20 minutes, at 0°C) through two successive 20% sucrose gradients in 1x buffered saline gelatin (BSG: 85%(w/v) NaCl, 0.03%(w/v) KH_2PO_4, 0.06%(w/v) Na_2HPO_4, 0.1%(w/v) gelatine). The minicells were removed by puncturing the side of the centrifuge tube and withdrawing the cells into a 10ml syringe. These were then washed in minimal medium and the optical density was adjusted to 0.6 at A_{650}. Minicells were pre-incubated in minimal medium to degrade long-lived mRNAs of chromosomally encoded genes, and pulse labelled with 12.5 µCi [^{35}S]-Methionine in the presence of Methionine assay medium (Difco) for 60 minutes. Following this, the cells were collected and resuspended in 1ml of nutrient broth and incubated at 37°C for 5 minutes. Minicells were pelleted by centrifugation for 3 minutes, resuspended in 100 µl of 1x sample buffer (Lugtenberg *et al.*, 1975) and solubilised by heating at 100°C for 5 minutes. The labelled proteins were analysed by SDS-PAGE and visualized by autoradiography.

2.32.2 Cell fractionation

After labelling the minicells, the fractionation was performed as a modification of the method of Osborn *et al.* (1972a, 1972b). The cells were pelleted and resuspended in 100µl of 20% sucrose, 30mM Tris-HCl pH 8.1 and chilled on ice. Cells were converted to sphaeroplasts with the addition of 10µl of 1mg/ml lysozyme in 0.1M EDTA pH 7.3 and incubating for 30 minutes on ice. After centrifugation at 10,000 r.p.m. for 10 minutes at 4°C, the supernatant was collected (periplasmic fraction). The cell pellet was frozen at -70°C for 30 minutes, thawed and dispersed vigorously in 300µl of 3mM EDTA pH 7.3

and lysed by 60 x 1 second bursts with a Branson Ultrasonifier. Intact cells and large cell debris were removed by low speed centrifugation (5,000 r.p.m, 5 minutes, 4°C) and the supernatant containing the membranes and cytoplasm was centrifuged at 35,000 r.p.m in a Beckman Ti50 rotor for 60 minutes at 4°C in a Beckman L8-80 ultracentrifuge. The supernatant (cytoplasmic fraction) was collected and the membrane pellet was resuspended in 25µl water and 75µl of 1.67% sarcosyl in 11.1mM Tris-HCl pH 7.6. After incubating at room temperature for 20 minutes the sarcosyl-insoluble material was recovered by centrifugation at 35,000 r.p.m in a 50Ti rotor for 90 minutes at 4°C. The sarcosyl-soluble supernatant contained the inner-membrane whereas the pellet contained the outer-membrane (Achtman *et al.*, 1979).

2.32.3 Prokaryotic *In vitro* transcription/translation

The prokaryotic DNA-directed translation kit was obtained from Amersham. Reactions were carried out according to the manufacturer's specifications, which involved reacting together the S30 cell extract, supercoiled plasmid DNA (CsCl purified at 2-5 µg/reaction), amino acids (excluding methionine), L-[^{35}S]-Methionine and the supplement solution (containing the nucleotides, tRNA, inorganic salts and an energy generating system) at 37°C for 1 hour. The reaction was continued for a further 5 minutes, after the addition of Methionine chase solution, and terminated by placing the reaction tubes in an ice bath. Samples were diluted 1:1 (v/v) with the loading buffer (0.08M Tris-HCl pH 6.8, 0.1M DTT, 2% (w/v) SDS, 10% (v/v) glycerol, 0.1 µg/ml bromophenol blue), heated to 100°C for 5 minutes and then analysed by SDS-PAGE and autoradiography.

2.32.4 SDS-Polyacrylamide Gel Electrophoresis of proteins

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on 15% polyacrylamide gels using a modification of the procedure of Lugtenberg *et al.* (1975), as described previously by Achtman *et al.* (1978). Samples were heated at 100°C for 5 minutes prior to loading. Gels were run at 150V for 3-4 hours or at a constant current of

10 mA for 16 hours. Proteins were stained by either, (1) shaking gently with the stain (10% methanol, 12% ethanol, 7.5% acetic acid and 0.0275% (w/v) coomassie brilliant blue R-250) for 15 to 45 minutes and destaining in a mix containing methanol (10%), ethanol (10%) and acetic acid (7.5%), with several changes at 15-30 minute intervals or, (2) by gentle agitation overnight at room temperature in 0.06% (w/v) coomassie brilliant blue G250 in 5% (v/v) perchloric acid and destaining in several changes of 5% (v/v) acetic acid, with gentle agitation, for 6-24 hr.

2.32.5 Autoradiography

Protein gels were dried on Whatmann 3MM chromatography paper at 60°C for 2 hours on a gel drier (BIO-RAD) before autoradiography. Kodak XR-100 film was used at room temperature for 1-7 days without the use of intensifying screens. With [^{32}P]-labelled samples, the gels were exposed to the X-ray film using intensifying screens for 1-72 hours at -70°C.

2.32.6 Preparation of CS3 antisera

Adults rabbits were obtained from the Central Animal House of the University of Adelaide for the production of antisera to CsCl purified native CS3 fimbriae or to denatured CS3 protein by subcutaneous immunization. The rabbit was immunized without adjuvant by subcutaneous injections on day 1, 10 and 17. The rabbits were bled by cardiac puncture under anaesthesia 10 days after the last immunization.

2.32.7 Crude pili preparation

An overnight culture (0.1ml) of the bacterial strain was plated on CFA agar plate containing the appropriate antibiotic and incubated at 37°C overnight. The cells were harvested into 0.75 ml of PBS, incubated at 56°C for 20 minutes and centrifuged in an Eppendorf centrifuge for 5 minutes. The supernatant was removed and diluted with an equal volume of 2 x SDS sample buffer (Lugtenberg *et al.*, 1975), containing 0.6% bromophenol blue, and electrophoresed on 15% SDS polyacrylamide gels.

2.32.8 Colony transfer and blotting with antiserum

A nitrocellulose disc (9 cm diameter) was placed on agar plates containing the patched colonies to be screened. After overnight growth of the colonies at 37°C the disc was peeled off very gently from the plate, air dried and incubated for 30 minutes in 5% (w/v) skim milk powder in TTBS (section 2.5.6) to block the non-specific protein binding sites. The cell debris was removed with a jet of saline (0.9% w/v NaCl). The antiserum was diluted 1:1000 in TTBS containing 0.02% (w/v) skim milk powder, added to the filters and incubated with gentle agitation at room temperature for 2-16 hours. Excess antibody was removed by shaking the nitrocellulose disc in TTBS three times for 10 minutes each. The disc was then gently agitated for 1-4 hours with goat anti-rabbit IgG, coupled with horseradish peroxidase (KPL), at a dilution of 1:5000 in TTBS plus 0.2% (w/v) skim milk powder, and washed four times at 5 minute intervals with TTBS, followed by two 5 minutes washes in TBS (20 mM Tris-HCl, 0.9%(w/v) NaCl). To detect the antigen-antibody complex, a peroxidase substrate (9.9 mg 4-Chloro-1-Napthol dissolved in 3.3 ml -20°C methanol, added to 16.5 ml TBS containing 15 ml hydrogen peroxide) was added and the disc incubated for 10-15 minutes with shaking.

2.32.9 Western Blotting

The procedure used was a modification of that described by Towbin *et al.* (1979). Samples were run on SDS-PAGE and transferred to nitrocellulose at 15-20mA overnight in a trans-blot cell using 25 mM Tris-HCl pH 8.3, 192 mM glycine and 5% (v/v) methanol as transfer buffer. The nitrocellulose sheet was then soaked for 30 minutes in 5% (w/v) skim milk powder in TTBS to block non-specific protein binding sites and incubated with the antiserum, diluted 1/1000 in TTBS, 0.02% (w/v) skim milk powder, at room temperature for 2-16 hours. Excess antibody was removed by washing the nitrocellulose sheet four times for 5 minutes each in TTBS with shaking. Bound antibody was detected by incubating the filter for 2-4 hours with gentle agitation in goat anti-rabbit IgG coupled with horseradish peroxidase at a dilution of 1/5000 in TTBS plus 0.02% skim milk powder. The filter was then washed four times (5 minutes intervals) with

TTBS, followed by two 5 minutes washes in TBS (section 2.5.5). The antigen-antibody complexes were then visualised using either of the following staining procedures:

2.32.9.a Using peroxidase

The substrate (20mg 4-Chloro-1-Napthol, dissolved in 7 ml of cold methanol and added to 33 ml of 1 x TBS containing 50µl hydrogen peroxide) was added and allowed to incubate for 10-15 minutes with shaking. The reaction was stopped by adding distilled water.

2.32.9.b Using Di-aminobenzidine (DAB)

Following the wash with TBS the nitrocellulose was pre-stained by adding, separately, DAB (7.5 mg / 7.5 ml PBS) and $CoSO_4$ (5 mg / 7.5 ml PBS) at room temperature. The stain was then added (7.5 ml of 1mg/ml DAB in PBS, 9 ml of Hydrogen peroxide and 5 mg / 7.5 ml of $CoSO_4$ in PBS) and incubated at room temperature without shaking. The nitrocellulose sheet was kept away from direct light during staining. The bands usually appeared within one minute and the reaction was stopped by adding PBS + Na-Azide solution.

2.32.10 Electronmicroscopy

2.32.10.a Preparation of protein A-Gold

Colloidal gold (10nm-15nm in diameter) was treated with dextran, activated with glutaraldehyde and then coupled to protein-A (Pharmacia), according to the method of Hicks and Molday (1984).

2.32.10.b Immuno-gold labelling of whole cells

Immunolabelling and electron microscopy of whole cells were carried out as follows: A 30µl of a bacterial suspension in phosphate buffered saline pH 7.2, containing 3% bovine serum albumin was placed on a sheet of parafilm. A formvar electron microscope grid, treated with poly-l-lysine, was then placed (plastic side down) on the surface of the drop for 2 minutes. Excess liquid was removed and the grid was transferred to drops of PBS+BSA placed on the same parafilm sheet. The grid was

then placed face down on a drop of a suitable dilution (1:50) of antiserum for 15 minutes. After thorough washing in PBS, the grids were placed on drops of protein A-gold (1:50 dilution in PBS+BSA) for 15 minutes. After washing further, the grids were negatively stained with a solution of 2% (w/v) uranyl acetate. Grids were examined in Philips EM300 transmission electron microscope, operated with an accelerating voltage of 80 KV.

2.33 M13 cloning and sequencing

The insert in pPM484 was isolated after digesting the plasmid DNA with *Hind*III, ligating into the *Hind*III site of M13mp19 and the two opposite orientations were selected. These were used in generating progressive deletions as described below.

2.33.1 Generation of Subclones for sequencing

Two approaches were used to generate the subclones for determining the sequence.

2.33.1.a DNaseI deletion method

Firstly, progressive deletions from either end of the insert were made by a slightly modified version of the DNase-I deletion method of Anderson (1981), as follows: Double stranded DNA of the M13 clones was digested with *Bam*HI and dephosphorylated. DNaseI stock solutions (1mg/ml in 0.1N HCl, stored at -80°C) was thawed on ice and serially diluted to 10^{-5} in 500mM BSA, 10 mM Tris-HCl pH 8.0 and 10 mM $MnCl_2$. The cut and dephosphorylated DNA was treated with 1µl from each of these DNAse dilutions for 15 minutes and run on 1% agarose gel to check the extent of digestion. The sample that showed between 10 and 30% digestion, indicated by a smear of DNA below the linear DNA band, was endfilled with Klenow in the presence of all four dNTPs and self ligated. Upon transfecting into JM101 this ligation mix produced several plaques, some of which were blue on the indicator plates. The white plaques were picked, inoculated into fresh JM101 and cultured for 5 hours. About 20µl from the supernatant was mixed with 5µl of 5 x SDS buffer (3% SDS, 250mM EDTA and 50mM Tris-HCl pH 8.0),

heated to 100°C for 5 minutes and run on 1% agarose gel along with single strand DNA templates from M13mp19 and one of the clones of the *Hin*dIII fragment. Deletion derivatives were selected based on their mobility in relation to these markers, such that the estimated size difference between two adjacent ones were approximately 200 bases and the total number of clones was enough to cover the entire fragment. The same procedure was repeated with the DNA from the clone in the opposite direction.

2.33.1.b Restriction fragment subcloning method

The *Hin*dIII fragment from pPM484 was digested separately with *Taq*I, *Sau3*AI and *Hpa*II, and end-labelled with Klenow and α-[^{32}P]-dATP. The individual fragments isolated from polyacrylamide gels were cloned into M13mp18 or mp19 RF DNA, which had been cleaved with either *Acc*I (for the *Taq*I and *Hpa*II fragments) or *Bam*HI (for the *Sau*3AI fragments).

2.33.2 Preparation of M13 replicative form (RF) DNA

Fresh 2 x YT broth (10 ml) was inoculated with 10 µl of an overnight culture of JM10/JM109/TG1 (in M13 minimal medium). A single plaque from the plate, picked up with a sterile Pasteur pipette, was inoculated into the culture and grown at 37°C with vigorous shaking for 6 hours. The culture was then added to 1 litre of 2 x YT broth and incubation continued overnight at 37°C with shaking. Replicative form (RF) DNA was prepared as described above for plasmid DNA purification, using a CsCl/Ethidium Bromide equilibrium gradient.

2.33.3 Purification plaques for sequencing

When the density of plaques on the plate was high, the plaques were purified by the following procedure:

1ml of 2xYT broth in a sterile 1.5ml eppendorf tube was inoculated with a single plaque picked up with a Pasteur pipette. This was incubated for 5minutes at room temperature and then 5µl was transferred into 5ml of fresh 2x YT broth. After mixing thoroughly 2.5 to 5µl of this diluted phage suspension was plated on M13 minimal plates

containing the indicator bacteria, BCIG and IPTG. The plates were incubated overnight at 37°C and single plaques isolated for further inoculations.

2.33.4 Screening M13 clones for inserts

Colourless plaques were picked from BCIG/IPTG plates with sterile toothpicks and added to 1 ml of 2xYT broth in microfuge tubes containing a 1:100 dilution of an overnight culture of JM101/TG1/JM109. The cultures were incubated for 5hr at 37°C and the cells were pelleted by centrifugation (10 minutes, Eppendorf). For rapid screening of inserts, 25µl of the supernatant was lysed by adding 5ml of 5 x lysis buffer (3% SDS, 50mM Tris-HCl pH 8.0, 1.5M NaCl and 250mM EDTA) and incubating at 65°C for 30 minutes. The lysed phage suspension was run on a 1% agarose gel along with clones not containing the insert as controls. By checking the differences in the mobility, clones with insert were identified. For further checking, the RF DNA suitable for restriction analysis were prepared from selected clones by the miniprep method and run on agarose gels after restricting with a suitable enzyme.

2.33.5 Single stranded phage preparation

Overnight cultures of JM101 or TG1 grown in M13 minimal medium were diluted in 2xYT broth and 2 ml each dispensed into screw capped polycarbonate tubes. After 15 minutes of incubation at 37°C with aeration the cultures were infected with M13 phage by picking fresh plaques using a sterile tooth pick and incubated at 37°C with aeration for 5 to 6 hours or overnight. The cells were sedimented for 15 minutes in an Eppendorf. The supernatant was transferred to clean tubes and re-centrifuged for further 10 minutes. About 1ml of supernatant was transferred to an eppendorf tube, and the phage particles were precipitated by mixing with 270µl of PEG solution (20% PEG and 2.5M NaCl) at room temperature for 15 minutes. The phage pellet was collected by centrifugation at room temperature for 5 minutes and, following another short spin for 10 seconds, the remainder of the PEG/NaCl supernatant was removed with a drawn out

Pasteur pipette. The phage pellet was resuspended in 200µl of TE. Phage stocks prepared in this way were stored at 4°C.

2.33.6 Single stranded DNA preparation

The phage suspension (200µl) was phenol extracted with 100µl of TE-saturated phenol at room temperature. From the aqueous phase, 150µl was carefully withdrawn and ethanol precipitated overnight at -20°C after adding 6µl of 3M sodium acetate pH 4.8 and 400µl of ethanol. Single stranded DNA was precipitated by centrifugation for 15 minutes in an Eppendorf centrifuge. DNA pellets were washed once with 1 ml of 95% (v/v) ethanol followed by centrifugation. After drying *in vacuo* the pellets were resuspended in 20µl of 1x TE and stored at -20°C until required.

2.33.7 Sequencing

Several variations of the dideoxy sequencing were used during the course of this project. These are described below:

2.33.7.1 Dideoxy sequencing protocol with Klenow fragment

2.33.7.1.a Annealing of primer and template

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

2.33.7.1.b Polymerization reaction

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

Prior to sequencing, reaction mixes were prepared by combining equal volumes of ddNTPs and dispensing 2µl each into eppendorf tubes. The annealed DNA-Primer was mixed with 2µl of lyophilised α-[^{32}P]-dCTP and 2µl of label supplement (16mM dCTP in 5mM Tris-HCl pH 8.0 and 0.1mM EDTA). From this, 2µl each was dispensed into four reaction tubes. The Klenow enzyme solution (1 unit/µl) was diluted just before use, and 2µl (approx. 0.5units) each was dispensed onto the side walls of the reaction tubes.

Table 2.3

Sequencing Reagents

a. dNTP solutions

Components	Composition			
	A°	C°	G°	T°
dATP	11 µM	153 µM	215 µM	215 µM
dGTP	215 µM	153 µM	16 µM	215 µM
dTTP	215 µM	153 µM	215 µM	16 µM

Symbols A°, C°, G°and T° refer to the respective mixes containing the rate limiting deoxy-ribo nucleotide triphosphate.

The sequencing reagents were prepared in 5mM Tris-HCl pH 8.0 and 0.1mM EDTA, by mixing the required amounts of dNTP solutions from 20mM stocks (in water), and stored at -20°C.

b. ddNTP solutions

Components	Composition
ddATP	0.5mM
ddCTP	0.1mM
ddGTP	0.3mM
ddTTP	0.8mM

The ddNTP solutions were made in water.

The sequencing reaction was commenced by a quick centrifugation to bring the enzyme solution down into the reaction mix.

After exactly 15 minutes at 37°C, 2µl of dNTP solution (0.25mM each of dATP, dCTP, dGTP and dTTP in 1xTE), containing 0.25 units of the Klenow enzyme, was added to each tube. The reaction was continued for a further 15 minutes and then terminated by adding 4µl of freshly prepared formamide loading buffer (95% deionised formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10mM EDTA and 10mM NaOH). Just before loading onto the gel the reaction mixes were heated to 100°C for 3 minutes, and 0.5µl each loaded onto 6% polyacrylamide denaturing gels. The gels were run at a constant 1500V until the bromophenol blue dye had migrated to about 2cm from the bottom of the gel.

2.33.7.1.c Fixing of gel and autoradiography

After removing the glass plate the gel was immediately washed with a solution containing 10% acetic acid and 20% ethanol. The washing was continued for about 15 minutes until all urea had been removed from the gel. The gel was then baked in a 110°C oven for 45 minutes and autoradiographed overnight at room temperature.

2.33.7.2 Dideoxy sequencing protocol with Sequenase

The dideoxy chain termination procedure of Sanger *et al.* (1977) was modified to encompass the use of Sequenase (modified T7 DNA Polymerase) in place of Klenow (Tabor and Richardson, 1987). All reagents were stored at -20°C. Two types of labelling and termination mixes were used, namely the dGTP mixes and the dITP mixes. The contents of the dGTP mixes are as follows:

Labelling mix :	7.5mM dGTP, dCTP and dTTP.
ddG termination mix :	80mM dNTP, 8mM ddGTP, 50mM NaCl.
ddA termination mix :	80mM dNTP, 8mM ddATP, 50mM NaCl.
ddC termination mix :	80mM dNTP, 8mM ddCTP, 50mM NaCl.
ddT termination mix :	80mM dNTP, 8mM ddTTP, 50mM NaCl.

The dITP mixes were used to reduce gel artefacts due to secondary structures in DNA synthesised in the sequencing reaction (Barnes *et al.*, 1983; Gough and Murray, 1983). The dITP mixes are as follows:

Labelling mix : 15mM dITP, 7.5mM dCTP and 7.5mM dTTP.

ddG termination mix : 160mM dITP, 80mM dATP, dCTP, dTTP, 1.6mM ddGTP, 50mM NaCl.

ddA termination mix : 160mM dITP, 80mM dATP, dCTP, dTTP, 8 mM ddATP, 50mM NaCl.

ddC termination mix : 160mM dITP, 80mM dATP, dCTP, dTTP, 8 mM ddCTP, 50mM NaCl.

ddT termination mix : 160mM dITP, 80mM dATP, dCTP, dTTP, 8 mM ddTTP, 50mM NaCl.

Normally, the labelling mix was diluted 1:5 with water to obtain the working concentration, however, to read long sequences in a single reaction, a dilution of 4:5 was used. The synthetic primer was annealed to the template by incubating 7µl of template (5-10 nM), 1µl of primer (500nM) and 2µl of 5 x sequenase buffer (200mM Tris-HCl pH7.5. 100mM $MgCl_2$, 250mM NaCl). The mixture was heated in a metal block at 65°C for 5-15 minutes and then the block containing the tubes was allowed to cool to room temperature. To the annealed mixture was added, 2µl of the appropriately diluted labelling mix, 1µl of 0.1MDTT, 0.5µl of [α-^{35}S]dATP (1000 Ci/mmol) and 2µl of diluted sequenase (1:8 dilution in 1x TE buffer). The reaction mixture was incubated for 5 minutes at room temperature, and 3.5µl each was aliquoted into four microfuge tubes, which have been pre-warmed to 37°C and labelled A, C, G and T, each containing 2.5µl of the corresponding termination mix. A quick centrifugation brought the reaction and termination mixes together to start the termination reaction. After 3 to 5 minutes at 37°C, 4µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05%

Xylene cyanol) was added to each of the reactions. Reaction mixes were heated to 100°C for 2minutes and 1.5µl each loaded onto the sequencing gel.

2.33.7.3 Double stranded DNA sequencing (CsCl DNA)

Plasmid DNA (2-4µg) was diluted to 32µl with water. The DNA was denatured by the addition of 4µl each of 2M NaOH (freshly prepared) and 2mM EDTA. The mixture was incubated for 5 minutes at room temperature. To this mix was added 16µl of 5M ammonium acetate pH 7.5 and 200µl of 95% ethanol, to precipitate the DNA. The supernatant was removed after centrifugation at 4°C for at least 30 minutes. The pellet was washed in 70% ethanol. The pellet was dried in vacuum and dissolved in a mixture containing 1µl of primer (500nM), 2µl of 10xTM and 7µl of water and heated to 37°C for 20-60 minutes. The labelling and termination reactions were done exactly as described for M13 single stranded template DNA (section 2.33.7.1b).

2.33.7.4 Double stranded DNA sequencing (miniprep DNA)

Plasmid DNA (2-4µg) from the 1ml preparation of 18µl/20µl volume in water was denatured by the addition of 2µl of 2M NaOH (freshly prepared) and 2mM EDTA mix. The mixture was incubated for 5 minutes at room temperature. To this mix, 8µl of 5M ammonium acetate pH 7.5 was added with 100µl of 95% ethanol to precipitate the DNA. The DNA was pelleted at 4°C for at least 30 minutes and washed in 70% ethanol. The pellet was dried in vacuum, dissolved in a mixture containing 1µl of primer (500nM), 2µl of 10 x TM and 7µl of water and heated to 37°C for 20-60 minutes. The labelling and termination reactions were run exactly as described for M13 single-stranded template DNA (see section 2.33.7.1b).

Sequence compilation was done on a Vax 11/750 computer using a modified version of the DNA sequence comparison program, DBCOMP (Staden, 1980) for the general compilation and analysis of the sequence. The codon usage searching program, GENE (Kalionis, *et al.*, 1986) was used to search the DNA for possible coding regions by comparing the codon usage against a standard codon-frequency table prepared from

known ETEC genes, using the method of Staden and McLachlan (1982). Various other comparisons were performed using the software package of the Genetics Computer Group (GCG; Devereux *et al.*, 1984). Sequence comparisons to the various nucleotide and protein databases (GenBank, EMBL, NBRF-PIR and SWISS-PROT) were by the FASTA program of Lipman and Pearson (1985). More recently, comparisons were also done using the BLAST algorithm, provided by the GenBank online services, to search the databases.

2.34 Site-directed oligonucleotide mutagenesis

Oligo-mutagenesis were performed by two different methods:

2.34.1 Method 1

Using the Amersham oligonucleotide-directed *in vitro* mutagenesis system version 2 (code RPN.1523). The procedure was followed according to the manufacturers' specification. Clones were confirmed by single stranded DNA sequencing.

2.34.2 Method 2

Single strand templates from M13 clones of the appropriate polarity were prepared as described earlier (section 2.33.6) and the kinased oligonucleotide carrying the desired change, as well as the universal sequencing primer, were annealed to it by incubating at 60°C for 1 hour. Synthesis of the complementary strand and ligation was achieved by incubating with Klenow, the four dNTPs (0.25mM each) and T4 DNA ligase for overnight. The ligated DNA was transfected into TG1 competent cells. The plaques were screened by a modified method of Benton and Davis (1977) as follows: The plates were chilled at 4°C for 30 minutes and a dry nitrocellulose disc was placed on the plates for 30minutes. The disc was then peeled off very gently, air dried and baked at 80°C for 2 hours *in vacuo*. The discs were hybridized overnight at 41°C with the 5' labelled oligonucleotide, in a hybridization solution containing 0.9M NaCl, 90mM Tris-HCl pH 7.5, 6 mM EDTA, 5 x Denhardt's solution, 100µg/ml Salmon sperm DNA and 0.5% Non-idet P-40. The nitrocellulose disc was washed twice in 6xSSC at room temperature,

followed by two washes in the TMACl (section 2.5.9) solution at 66 to 69°C, depending on the length of the oligonucleotide. The actual temperature of the solution was monitored by placing a thermometer in it. After washing, the discs were air-dried, wrapped in a plastic and exposed to X-ray film overnight at -80°C with intensifying screen. Positive clones were purified and analysed either by sequencing or, in cases where a new restriction site was created by the mutation, by restriction analysis.

CHAPTER 3

Sequence Analysis Of The DNA Encoding The CS3 Gene Cluster

CHAPTER 3

Sequence Analysis Of The DNA Encoding The CS3 Gene Cluster

3.1 Introduction

Shotgun cloning of *Hin*dIII digested plasmid DNA from the ETEC strain PB176 resulted in the isolation of a number of clones capable of synthesising CS3 pili in *E. coli* K-12 (Manning *et al.*, 1985). All carried a common 4.6 kb fragment, termed the minimal fragment, which, contained the essential plasmid-encoded genes for CS3 pilus biosynthesis. Minicell labelling, using the the *E. coli* strain DS410 (Dougan and Sherratt, 1977), showed that pPM474, the plasmid that carried the minimal fragment and an additional 0.71 kb of contiguous DNA, coded for at least seven polypeptides with relative molecular weights (MW) of 97000, 94000, 58000, 46500, 27000, 18000 and 16600, respectively. Deletion derivatives of pPM474 which did not form CS3 pili, suggested that most, if not all of these polypeptides, were involved in its biosynthesis.

Based on the translation pattern observed with pPM474 and its deletion derivatives, Manning *et al.* (1985) assumed that the 94 kDa, 46.5 kDa and 16.6 kDa polypeptides were derived from the 97 kDa, 58 kDa, and 18 kDa proteins, respectively. An overlapping gene arrangement was proposed, as the length of the minimal fragment (4.6 kb) was too short to encode these three precursors, together with the 27 kDa polypeptide, as independent transcription and/or translation units. It appeared that only the 27 kDa polypeptide and the CS3 subunit genes could be non-overlapping with any other gene on the minimal fragment.

The plasmid, pPM484, was similar to pPM474 in its ability to direct CS3 biosynthesis, but it carried only the minimal *Hin*dIII fragment (*see* Figure 6 in Manning *et al.*, 1985). This plasmid was chosen for the present study, which is aimed at providing an understanding of the molecular organisation of the CS3 operon. Evidence is presented to substantiate the earlier prediction that all essential genes for CS3 biosynthesis,

encoded on the parental plasmid in the wild type ETEC strain, PB176, are present on pPM484. To identify these genes, and to assign possible roles to their protein products, the nucleotide sequence of the minimal fragment was determined. The sequence and its computer analyses are described in this chapter.

3.2 Results

3.2.1 pPM484 encodes all genes required for CS3 pilus synthesis in *E. coli* K-12

The ability of *E. coli* K-12 strains to synthesise CS3 was confirmed by preparing crude pili from cultures of DS410 or C600 that have been transformed with pPM484 or pPM474. Pili were prepared from cells grown overnight as a thick layer on CFA plates. Western blotting of this material, using antiserum raised against CsCl-purified CS3 pili from PB176, revealed a single band of about 15 kDa (Figure 3.1, lanes *c* to *f*). In contrast, crude pili from PB176 showed two forms (Figure 3.1, lane *g*) - a major form that migrated slightly faster than the 15 kDa band, seen in *E. coli* K-12, and a minor form that co-migrated with it. The pili from *E. coli* K-12 strains showed a total lack of the shorter, major band. One possible explanation for this difference is that two sites for CS3 signal peptide processing exist, and that only one is being used in *E. coli* K-12, whereas both are being used to different extents in ETEC. Evidence for this interpretation will be presented later (*see* section 4.2.1.8 in Chapter 4). Judging from the relative intensities of the CS3 bands, as compared to the non-specific bands in the same lanes, DS410 produces a relatively higher level of pili than C600. Immunogold labelling, followed by electron microscopy of DS410 (pPM484) revealed the presence of CS3 fimbriae on the cell surface (Figure 3.2 b). The amount of pili was much lower than that found on the ETEC cell surface (Figure 3.2 a). Cells carrying pBR322, in contrast, were devoid of pili (Figure 3.2 c).

Labelling of the minicells, carrying pPM484, with ^{35}S-Met revealed that the pattern of protein bands (Figure 3.3, lane *a*) was identical to that obtained with pPM474

Figure 3.1 Western blot of crude pili from pPM474 vs pPM484

Comparison of pilus biosynthesis mediated by pPM474 and pPM484. Crude pili from DS410 and C600, carrying either pPM474 or pPM484, were run in SDS on 15% PAGE, transferred to nitrocellulose paper and probed with CS3 antiserum. Strains carrying pBR322 and the wild type ETEC strain, PB176, were used as controls. The two mature forms of CS3 pili subunits in PB176 are indicated.

Lanes:

a - DS410, carrying pBR322

b - C600, carrying pBR322

c - DS410, carrying pPM484

d - DS410, carrying pPM474

e - C600, carrying pPM484

f - C600, carrying pPM474

g - PB176.

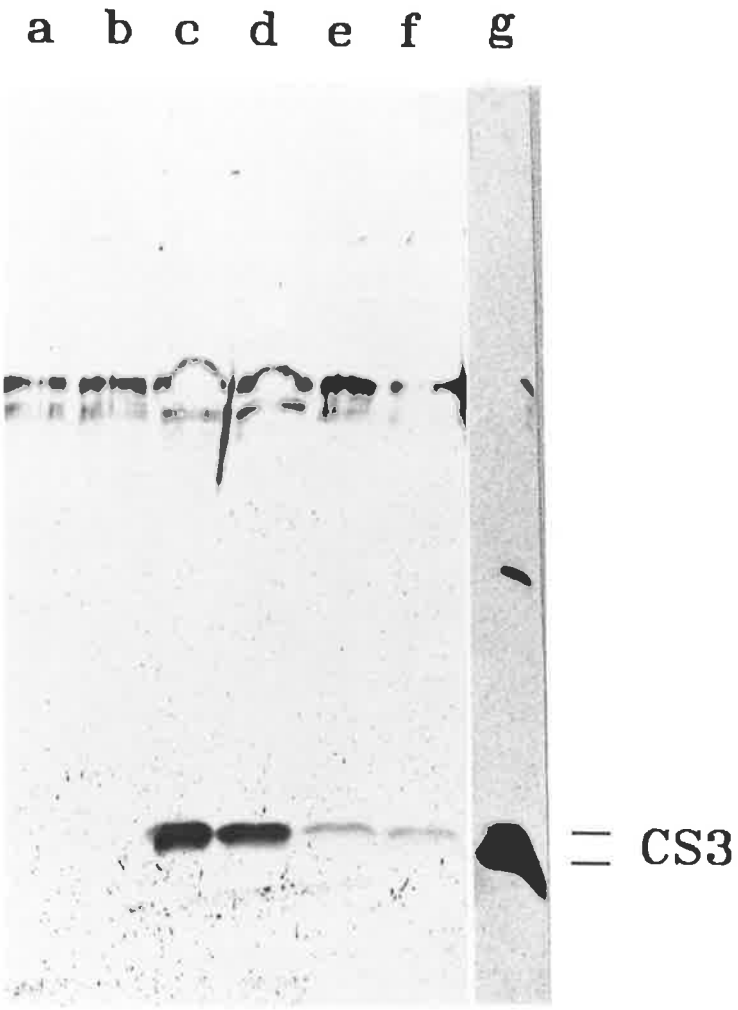

a b c d e f g
CS3

Figure 3.2 Electron micrographs showing cell-surface CS3 pili

An ETEC cell, and *E. coli* K-12 minicells carrying pPM484 or pBR322 were labelled with rabbit anti- CS3 pilin antibody followed by protein-A colloidal gold complexes.

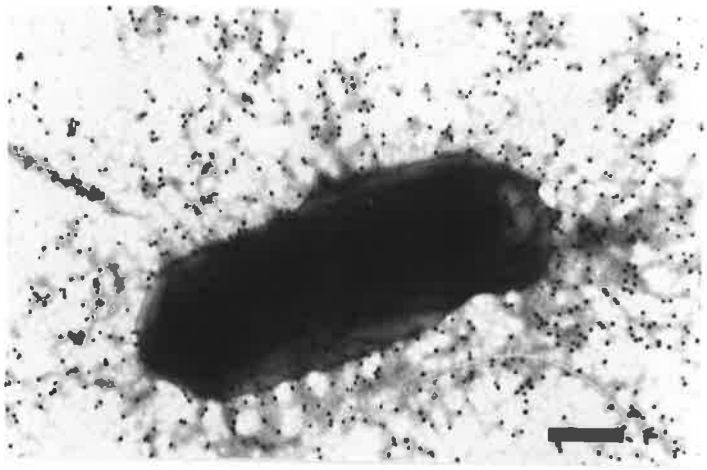
a - ETEC (PB176) cell

b - *E. coli* K-12 minicell, carrying pPM484

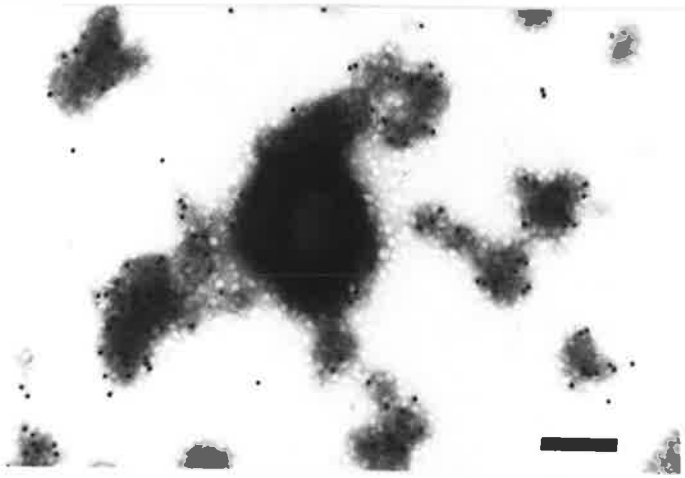
c - *E. coli* K-12 minicell, carrying pBR322

The photographs are taken at different magnifications. Thus, the horizontal bar equals 0.37μM in **a**, 0.30μM in **b**, and 0.37μM in **c**.

a



b



c

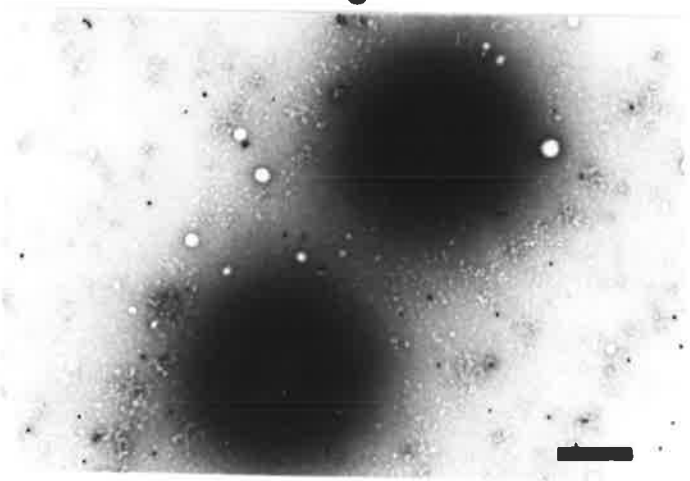


Figure 3.3 Minicell translation of pPM484 and pBR322

Minicells carrying either pPM484 or pBR322 were labelled with ^{35}S-Met, lysed and electrophoresed in SDS on 15% PAGE.

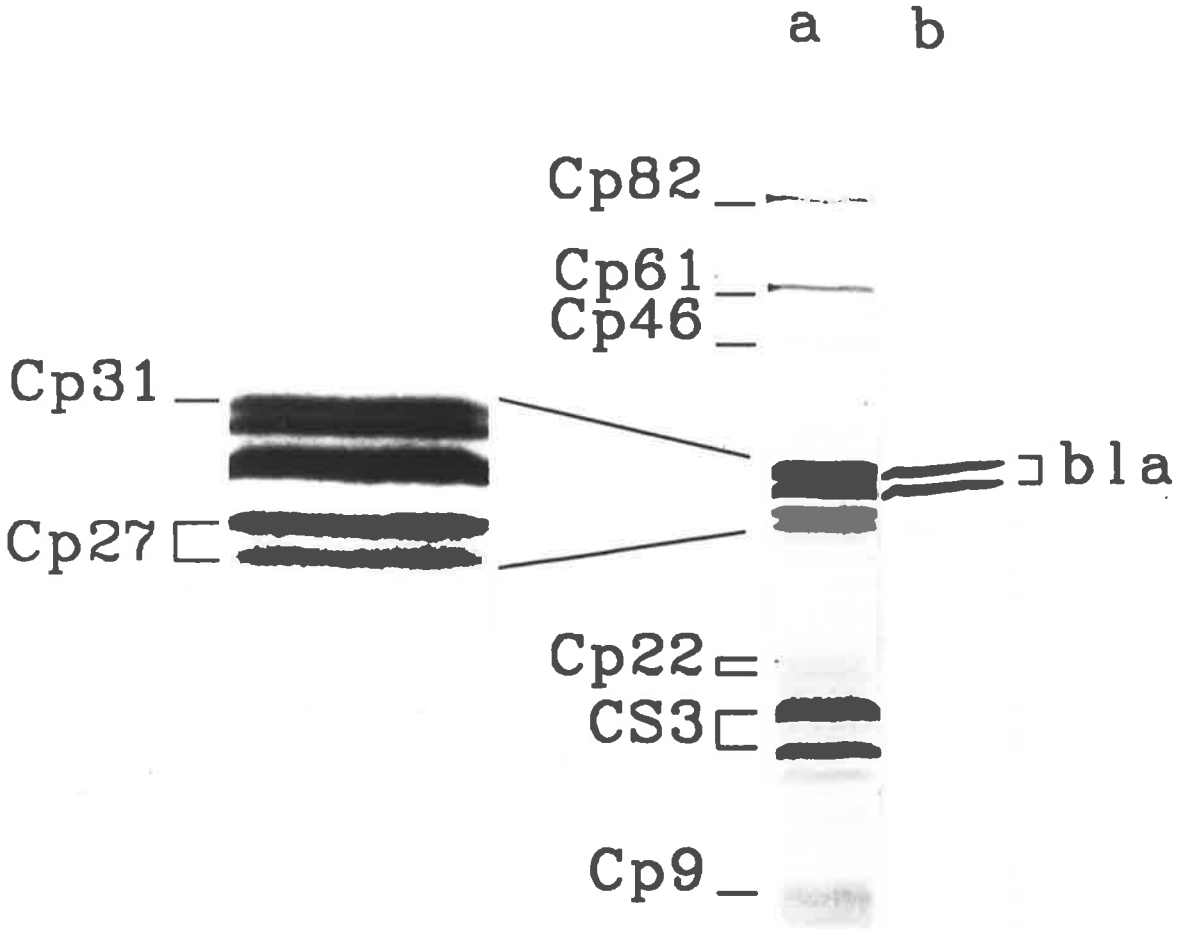
Lanes:

a - pPM484

b - pBR322

Molecular weight markers are:- phosphorylase B (97.4 kDa), bovine serum albumin (68 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.3 kDa). The approximate sizes (kDa) of pPM484-specific protein bands are indicated.

To be confident that the 9 kDa protein is true size you need a standard of this size or smaller.


a
b
Cp82
Cp61
Cp46
Cp31
Cp27
bla
Cp22
CS3
Cp9

by Manning *et al.* (1985). The bands designated as 97 kDa, 58 kDa, and 46.5 kDa were thin, but consistently present, whereas more intense bands were observed for the 27 kDa, 18 kDa and 16.6 kDa polypeptides. Along with these expected proteins, three others with sizes 31 kDa, 22 kDa and 9 kDa were also observed. Though not described by Manning *et al.* (1985), a re-investigation revealed that they were also being produced by pPM474. Presumably, these bands were overlooked in earlier studies owing to the masking effect of the β-lactamase on the 31 kDa protein, poor expression of the 22 kDa protein and small size of the 9 kDa protein.

3.2.2 The nucleotide sequence of pPM484

The insert from pPM484 was subcloned into M13 vectors and sequenced as described in Chapter 2. The complete sequence (4747 bp) and the sequencing strategy are shown in Figure 3.4a and b, respectively. Numbering is in the 5' to 3' direction, starting from the first base of the *Hin*dIII site, and has the same polarity as the mRNA. All regions in this sequence have been determined from more than one clone and in both directions. No disparity was observed between the sequences determined from independent clones. Areas that showed "compressions" were resolved by accepting the sequence of the strand not showing such compression. There was no region that could not be read unambiguously from one or the other strand.

3.2.3 Analysis of the nucleotide sequence

The nucleotide sequence was searched for the presence of various motifs that would help to identify the potential coding regions of the polypeptides observed in minicell translations. These included the open reading frames (ORFs), ribosome binding sites (RBS), promoters and terminators. In addition, the sequence was also analysed for its coding probability by comparing the codon usage patterns against tables based either on the *E. coli* K-12 genes or on the sequenced pilus biosynthesis genes from ETEC and other pathogenic *E. coli* strains. A modified computer algorithm based on the Staden and McLachlan (1982) method for predicting coding probability, by comparing the three

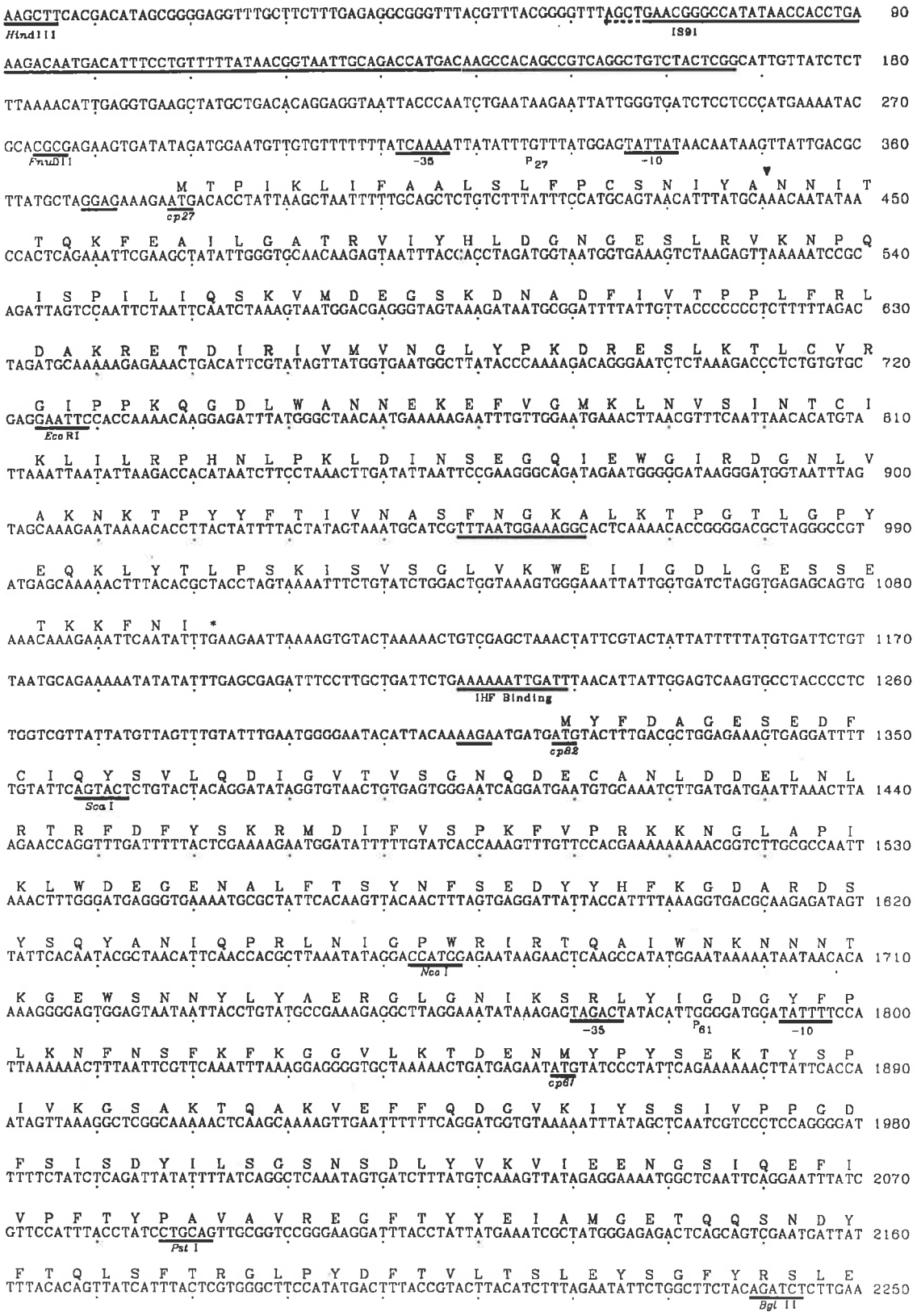
Figure 3.4 a. Nucleotide sequence of the *Hin*dIII fragment, encoding the CS3 pilus biosynthesis genes

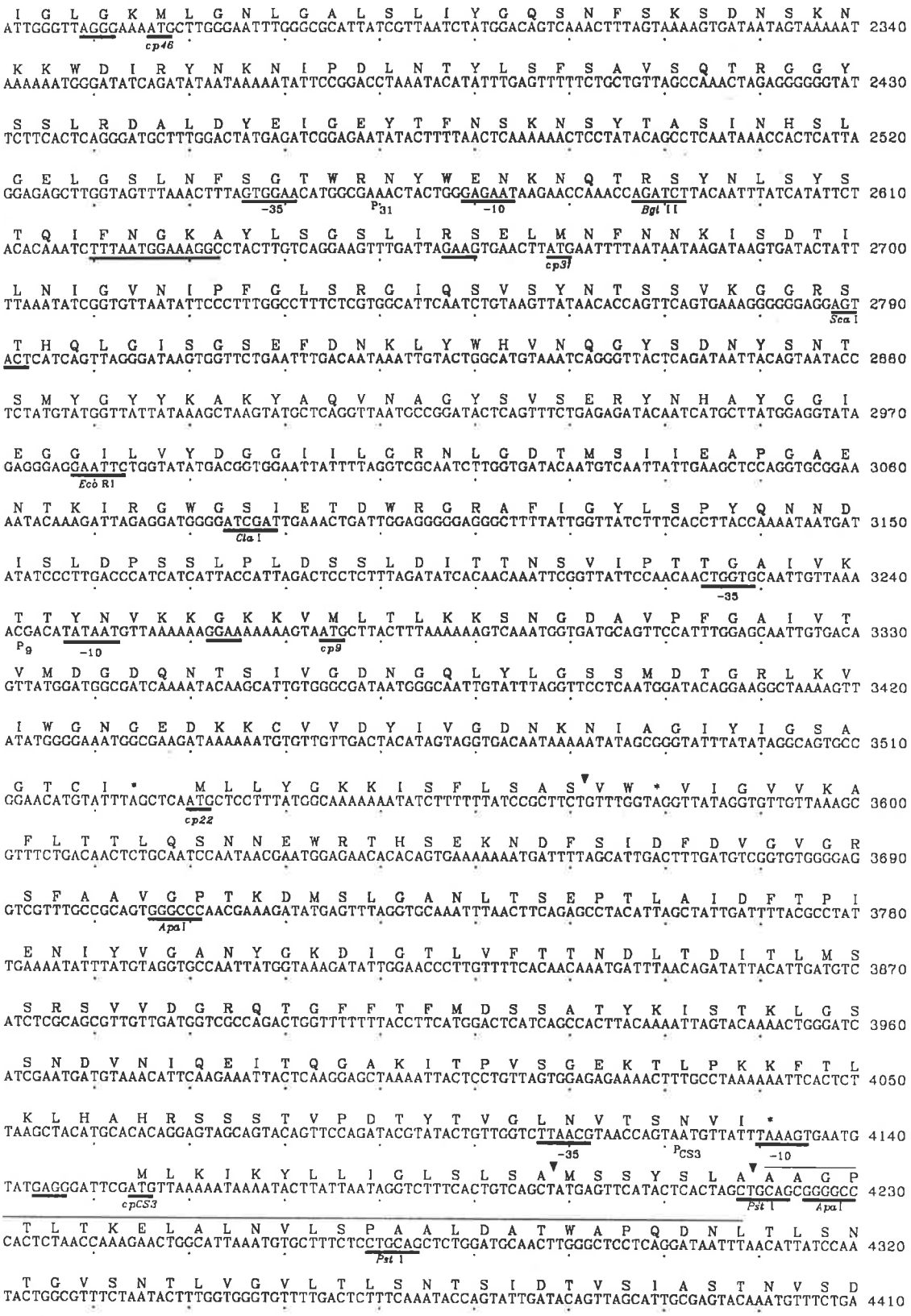
Nucleotide sequence of the *Hin*dIII fragment cloned in pPM484. The sequence reads in the 5' to 3' direction. The -35 and -10 regions of the promoters are underlined. Also underlined are the relevant restiction sites, the IHF binding site (AA---RTTGATR; Nash, 1981; Friedman, 1988) putative RBSs and initiation codons of the genes, the two 14 bp direct repeats (nt 948 and 2620) and the region homologous to the IS*91* element (section 3.2.6.1a). The deduced amino acid sequences of the genes are are written above the DNA sequence, and the sequenced N-terminal amino acids of CS3 are over-lined. Signal peptide cleavage sites of the polypeptides are marked by inverted triangles. A putative stem-loop structure after the CS3 coding region is over-lined.

b. Sequencing strategy:

Solid line represent the entire sequence, with short vertical bars marking the sequence positions. Thin horizontal lines represent the gel readings from individual sequencing experiments, the lengths of which correspond to the actual number of bases read. Gel readings in the 5' to 3' direction are marked above, whereas those in the opposite direction are given below the line representing the completed sequence. (The arrows represent the directions of gel readings).

a


AAGCTTCACGACATAGCGGGGAGGTTTGCTTCTTTGAGAGGCGGGTTTACGTTTACGGGGTTTAGCTGAACGGGCCATATAACCACCTGA 90
Hind III
IS91
AAGACAATGACATTTCCTGTTTTTATAACGGTAATTGCAGACCATGACAAGCCACAGCCGTCAGGCTGTCTACTCGGCATTGTTATCTCT 180
TTAAAACATTGAGGTGAAGCTATGCTGACACAGGAGGTAATTACCCAATCTGAATAAGAATTATTGGGTGATCTCCTCCCATGAAAATAC 270
GCACGCGAGAAGTGATATAGATGCGAATGTTGTGTTTTTTTTATCAAAATTATATTTGTTTATGGAGTATTATAACAATAAGTTATTGACGC 360
FnuDII
-35
P27
-10
M T P I K L I F A A L S L F P C S N I Y A N N I T
TTATGCTAGGAGAAAGAATGACACCTATTAAGCTAATTTTTGCAGCTCTGTCTTTATTTCCATGCAGTAACATTTATGCAAACAATATAA 450
cp27
T Q K F E A I L G A T R V I Y H L D G N G E S L R V K N P Q
CCACTCAGAAATTCGAAGCTATATTGGGTGCAACAAGAGTAATTTACCACCTAGATGGTAATGGTGAAAGTCTAAGAGTTAAAAATCCGC 540
I S P I L I Q S K V M D E G S K D N A D F I V T P P L F R L
AGATTAGTCCAATTCTAATTCAATCTAAAGTAATGGACGAGGGTAGTAAAGATAATGCGGATTTTATTGTTACCCCCCCTCTTTTTAGAC 630
D A K R E T D I R I V M V N G L Y P K D R E S L K T L C V R
TAGATGCAAAAAGAGAAACTGACATTCGTATAGTTATGGTGAATGGCTTATACCCAAAAGACAGGGAATCTCTAAAGACCCTCTGTGTGC 720
G I P P K Q G D L W A N N E K E F V G M K L N V S I N T C I
GAGGAATTCCACCAAAACAAGGAGATTTATGGGCTAACAATGAAAAAGAATTTGTTGGAATGAAACTTAACGTTTCAATTAACACATGTA 810
Eco RI
K L I L R P H N L P K L D I N S E G Q I E W G I R D G N L V
TTAAATTAATATTAAGACCACATAATCTTCCTAAACTTGATATTAATTCCGAAGGGCAGATAGAATGGGGGATAAGGGATGGTAATTTAG 900
A K N K T P Y Y F T I V N A S F N G K A L K T P G T L G P Y
TAGCAAAGAATAAAACACCTTACTATTTTACTATAGTAAATGCATCGTTTAATGGAAAGGCACTCAAAACACCGGGGACGCTAGGGCCGT 990
E Q K L Y T L P S K I S V S G L V K W E I I G D L G E S S E
ATGAGCAAAAACTTTACACGCTACCTAGTAAAATTTCTGTATCTGGACTGGTAAAGTGGGAAATTATTGGTGATCTAGGTGAGAGCAGTG 1080
T K K F N I *
AAACAAAGAAATTCAATATTTGAAGAATTAAAAGTGTACTAAAAACTGTCGAGCTAAACTATTCGTACTATTATTTTTATGTGATTCTGT 1170
TAATGCAGAAAAATATATATTTGAGCGAGATTTCCTTGCTGATTCTGAAAAAATTGATTTAACATTATTGGAGTCAAGTGCCTACCCCTC 1260
IHF Binding
M Y F D A G E S E D F
TGGTCGTTATTATGTTAGTTTGTATTTGAATGGGGAATACATTACAAAAGAATGATGATGTACTTTGACGCTGGAGAAAGTGAGGATTTT 1350
cp82
C I Q Y S V L Q D I G V T V S G N Q D E C A N L D D E L N L
TGTATTCAGTACTCTGTACTACAGGATATAGGTGTAACTGTGAGTGGGAATCAGGATGAATGTGCAAATCTTGATGATGAATTAAACTTA 1440
Sca I
R T R F D F Y S K R M D I F V S P K F V P R K K N G L A P I
AGAACCAGGTTTGATTTTTACTCGAAAAGAATGGATATTTTTGTATCACCAAAGTTTGTTCCACGAAAAAAAAACGGTCTTGCGCCAATT 1530
K L W D E G E N A L F T S Y N F S E D Y Y H F K G D A R D S
AAACTTTGGGATGAGGGTGAAAATGCGCTATTCACAAGTTACAACTTTAGTGAGGATTATTACCATTTTAAAGGTGACGCAAGAGATAGT 1620
Y S Q Y A N I Q P R L N I G P W R I R T Q A I W N K N N N T
TATTCACAATACGCTAACATTCAACCACGCTTAAATATAGGACCATGGAGAATAAGAACTCAAGCCATATGGAATAAAAATAATAACACA 1710
Nco I
K G E W S N N Y L Y A E R G L G N I K S R L Y I G D G Y F P
AAAGGGGAGTGGAGTAATAATTACCTGTATGCCGAAAGAGGCTTAGGAAATATAAAGAGTAGACTATACATTGGGGATGGATATTTTCCA 1800
-35
P61
-10
L K N F N S F K F K G G V L K T D E N M Y P Y S E K T Y S P
TTAAAAAACTTTAATTCGTTCAAATTTAAAGGAGGGGTGCTAAAAACTGATGAGAATATGTATCCCTATTCAGAAAAAACTTATTCACCA 1890
cp87
I V K G S A K T Q A K V E F F Q D G V K I Y S S I V P P G D
ATAGTTAAAGGCTCGGCAAAAACTCAAGCAAAAGTTGAATTTTTTCAGGATGGTGTAAAAATTTATAGCTCAATCGTCCCTCCAGGGGAT 1980
F S I S D Y I L S G S N S D L Y V K V I E E N G S I Q E F I
TTTTCTATCTCAGATTATATTTTATCAGGCTCAAATAGTGATCTTTATGTCAAAGTTATAGAGGAAAATGGCTCAATTCAGGAATTTATC 2070
V P F T Y P A V A V R E G F T Y Y E I A M G E T Q Q S N D Y
GTTCCATTTACCTATCCTGCAGTTGCGGTCCGGGAAGGATTTACCTATTATGAAATCGCTATGGGAGAGACTCAGCAGTCGAATGATTAT 2160
Pst I
F T Q L S F T R G L P Y D F T V L T S L E Y S G F Y R S L E
TTTACACAGTTATCATTTACTCGTGGGCTTCCATATGACTTTACCGTACTTACATCTTTAGAATATTCTGGCTTCTACAGATCTCTTGAA 2250
Bgl II

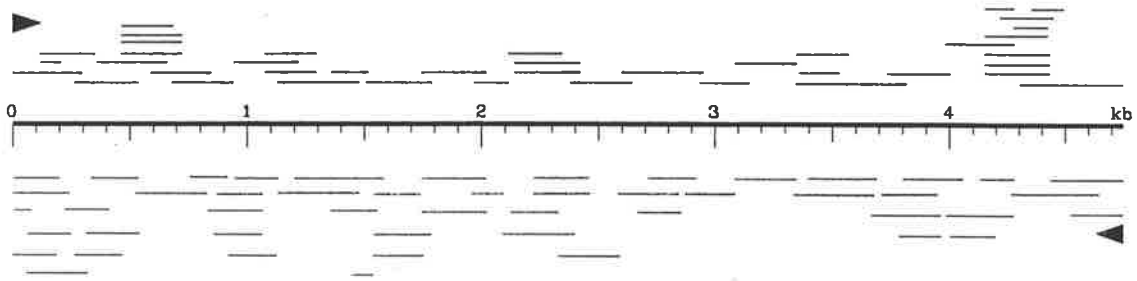

I G L G K M L G N L G A L S L I Y G Q S N F S K S D N S K N
ATTGGGTTAGGGAAAATGCTTGGGAATTTGGGCGCATTATCGTTAATCTATGGACAGTCAAACTTTAGTAAAAGTGATAATAGTAAAAAT 2340
cp46
K K W D I R Y N K N I P D L N T Y L S F S A V S Q T R G G Y
AAAAAATGGGATATCAGATATAATAAAAATATTCCGGACCTAAATACATATTTGAGTTTTTCTGCTGTTAGCCAAACTAGAGGGGGGTAT 2430
S S L R D A L D Y E I G E Y T F N S K N S Y T A S I N H S L
TCTTCACTCAGGGATGCTTTGGACTATGAGATCGGAGAATATACTTTTAACTCAAAAAACTCCTATACAGCCTCAATAAACCACTCATTA 2520
G E L G S L N F S G T W R N Y W E N K N Q T R S Y N L S Y S
GGAGAGCTTGGTAGTTTAAACTTTAGTGGAACATGGCGAAACTACTGGGAGAATAAGAACCAAACCAGATCTTACAATTTATCATATTCT 2610
-35
P31
-10
Bgl II
T Q I F N G K A Y L S G S L I R S E L M N F N N K I S D T I
ACACAAATCTTTAATGGAAAGGCCTACTTGTCAGGAAGTTTGATTAGAAGTGAACTTATGAATTTTAATAATAAGATAAGTGATACTATT 2700
cp3I
L N I G V N I P F G L S R G I Q S V S Y N T S S V K G G R S
TTAAATATCGGTGTTAATATTCCCTTTGGCCTTTCTCGTGGCATTCAATCTGTAAGTTATAACACCAGTTCAGTGAAAGGGGGGAGGAGT 2790
Sca I
T H Q L G I S G S E F D N K L Y W H V N Q G Y S D N Y S N T
ACTCATCAGTTAGGGATAAGTGGTTCTGAATTTGACAATAAATTGTACTGGCATGTAAATCAGGGTTACTCAGATAATTACAGTAATACC 2880
S M Y G Y Y K A K Y A Q V N A G Y S V S E R Y N H A Y G G I
TCTATGTATGGTTATTATAAAGCTAAGTATGCTCAGGTTAATGCCGGATACTCAGTTTCTGAGAGATACAATCATGCTTATGGAGGTATA 2970
E G G I L V Y D G G I I L G R N L G D T M S I I E A P G A E
GAGGGAGGAATTCTGGTATATGACGGTGGAATTATTTTAGGTCGCAATCTTGGTGATACAATGTCAATTATTGAAGCTCCAGGTGCGGAA 3060
Eco RI
N T K I R G W G S I E T D W R G R A F I G Y L S P Y Q N N D
AATACAAAGATTAGAGGATGGGGATCGATTGAAACTGATTGGAGGGGGAGGGCTTTTATTGGTTATCTTTCACCTTACCAAAATAATGAT 3150
Cla I
I S L D P S S L P L D S S L D I T T N S V I P T T G A I V K
ATATCCCTTGACCCATCATCATTACCATTAGACTCCTCTTTAGATATCACAACAAATTCGGTTATTCCAACAACTGGTGCAATTGTTAAA 3240
-35
T T Y N V K K G G K K V M L T L K K S N G D A V P F G A I V T
ACGACATATAATGTTAAAAAAGGAAAAAAAGTAATGCTTACTTTAAAAAAGTCAAATGGTGATGCAGTTCCATTTGGAGCAATTGTGACA 3330
P9
-10
cp9
V M D G D Q N T S I V G D N G Q L Y L G S S M D T G R L K V
GTTATGGATGGCGATCAAAATACAAGCATTGTGGGCGATAATGGGCAATTGTATTTAGGTTCCTCAATGGATACAGGAAGGCTAAAAGTT 3420
I W G N G E D K K C V V D Y I V G D N K N I A G I Y I G S A
ATATGGGGAAATGGCGAAGATAAAAAATGTGTTGTTGACTACATAGTAGGTGACAATAAAAATATAGCGGGTATTTATATAGGCAGTGCC 3510
G T C I * M L L Y G K K I S F L S A S V W * V I G V V K A
GGAACATGTATTTAGCTCAATGCTCCTTTATGGCAAAAAAATATCTTTTTTATCCGCTTCTGTTTGGTAGGTTATAGGTGTTGTTAAAGC 3600
cp22
F L T T L Q S N N E W R T H S E K N D F S I D F D V G V G R
GTTTCTGACAACTCTGCAATCCAATAACGAATGGAGAACACACAGTGAAAAAAATGATTTTAGCATTGACTTTGATGTCGGTGTGGGGAG 3690
S F A A V G P T K D M S L G A N L T S E P T L A I D F T P I
GTCGTTTGCCGCAGTGGGCCCAACGAAAGATATGAGTTTAGGTGCAAATTTAACTTCAGAGCCTACATTAGCTATTGATTTTACGCCTAT 3780
Apa I
E N I Y V G A N Y G K D I G T L V F T T N D L T D I T L M S
TGAAAATATTTATGTAGGTGCCAATTATGGTAAAGATATTGGAACCCTTGTTTTCACAACAAATGATTTAACAGATATTACATTGATGTC 3870
S R S V V D G R Q T G F F T F M D S S A T Y K I S T K L G S
ATCTCGCAGCGTTGTTGATGGTCGCCAGACTGGTTTTTTTACCTTCATGGACTCATCAGCCACTTACAAAATTAGTACAAAACTGGGATC 3960
S N D V N I Q E I T Q G A K I T P V S G E K T L P K K F T L
ATCGAATGATGTAAACATTCAAGAAATTACTCAAGGAGCTAAAATTACTCCTGTTAGTGGAGAGAAAACTTTGCCTAAAAAATTCACTCT 4050
K L H A H R S S S T V P D T Y T V G L N V T S N V I *
TAAGCTACATGCACACAGGAGTAGCAGTACAGTTCCAGATACGTATACTGTTGGTCTTAACGTAACCAGTAATGTTATTTAAAGTGAATG 4140
-35
PCS3
-10
M L K I K Y L L I G L S L S A M S S Y S L A A A G P
TATGAGGGATTCGATGTTAAAAATAAAATACTTATTAATAGGTCTTTCACTGTCAGCTATGAGTTCATACTCACTAGGCTGCAGCGGGGCC 4230
cpCS3
Pst I
Apa I
T L T K E L A L N V L S P A A L D A T W A P Q D N L T L S N
CACTCTAACCAAAGAACTGGCATTAAATGTGCTTTCTCCTGCAGCTCTGGATGCAACTTGGGCTCCTCAGGATAATTTAACATTATCCAA 4320
Pst I
T G V S N T L V G V L T L S N T S I D T V S I A S T N V S D
TACTGGCGTTTCTAATACTTTGGTGGGTGTTTTGACTCTTTCAAATACCAGTATTGATACAGTTAGCATTGCGAGTACAAATGTTTCTGA 4410

```
  T  S  K  N  G  T  V  T  F  A  H  E  T  N  N  S  A  S  F  A  T  T  I  S  T  D  N  A  N  I
TACATCTAAGAATGGTACAGTAACTTTTGCACATGAGACAAATAACTCTGCTAGCTTTGCCACCACCATTTCAACAGATAATGCCAACAT 4500

  T  L  D  K  N  A  G  N  T  I  V  K  T  T  N  G  S  Q  L  P  T  N  L  P  L  K  F  I  T  T
TACGTTGGATAAAAATGCTGGAAATACGATTGTTAAAACTACAAATGGGAGTCAGTTGCCAACTAATTTACCACTTAAGTTTATTACCAC 4590

  E  G  N  E  H  L  V  S  G  N  Y  R  A  N  I  T  I  T  S  T  I  K  *
TGAAGGTAACGAACATTTAGTTTCAGGTAATTACCGTGCAAATATAACAATTACTTCGACAATTAAATAATTATATAATAGACGTAGCCT 4680
```

T_{CS3}

```
TCGAAATAAAGGCTACGTTGCTATCTTTATGTTTGTGATTTATAGGCATCATTAAATAGTCAAGCTT                    4747
                                                                 HindIII
```

b



reading frames against the codon frequency tables, is shown as a plot in Figure 3.5. Whenever a reference to the sequence position of a particular signal is made, the convention of referring to the first nucleotide (nt) in the signal is followed.

3.2.4 Definition and summary of ORFs

An ORF is generally defined as a stretch of DNA not interrupted by stop codons. However, in the context of this thesis, an ORF has been defined as a stretch of DNA that codes for a string of amino acids starting with an AUG or GUG initiation codon. The initiation codons that defined the start of ORFs were not necessarily the first of their kind in an uninterrupted stretch of amino acids. In many cases, ORFs were arbitrarily defined, based on the size of the potential protein products they encoded. This re-defining has been made to accommodate the possibility that one or more of the polypeptides seen in minicell labelling could have inframe initiations within the reading frame of a larger polypeptide. In the instance of defining the ORF of one of the genes, *cp22*, a UAG triplet was assumed not to function as a termination codon. Reasons for this assumption are explained later in this chapter (section 3.2.4.3). Presence of suitable RBSs (Stormo *et al.*, 1982) in front of the initiation codons were also searched for, but were not deemed critical, as several proven genes in *E. coli* K-12 lacked clear RBSs (Gold *et al.*, 1981). It has been suggested that an RBS is not absolutely essential for initiation, instead is required for efficient translation (Calogero *et al.*, 1988).

Essentially, three criteria were used in the initial search of ORFs. Firstly, stretches of at least 75 amino acids (aa) from an initiation codon (AUG, GUG or UUG) were searched for. Secondly, presence of a suitable RBS (Stormo *et al.*, 1982; Munson *et al.*, 1984; Shinedling *et al.*, 1987; Hartz *et al.*, 1991) in the region preceding these potential initiation codons was examined. Finally, the codon usage pattern was analysed to gain an insight into the potential of them being protein coding regions. The presumptive genes thus identified were given the designation *cpnn* where *cp* stands for CS3 protein, and *nn* is a number approximating the predicted MW (x 1000; e.g. *cp82* for the ORF with MW

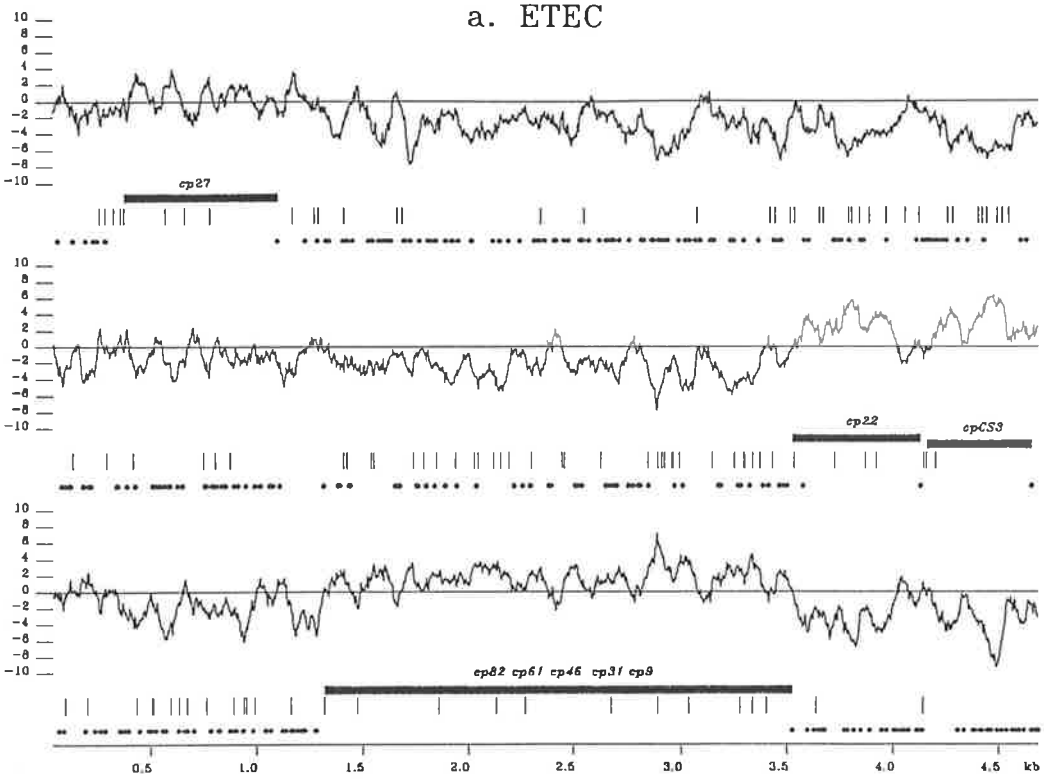
Figure 3.5 Codon usage plot of the DNA sequence

The codon usage in each of the three reading frames, on the coding strand alone, was calculated by a modified computer algorithm based on the method of Staden and McLachlan (1982). The probability of a triplet being in a protein coding region was calculated over a window of 25 codons, by comparing against the respective codon usage tables as well as against the two alternative reading frames. Probable coding regions generally show stretches of positive values uninterrupted by stop codons (marked by dots). Short vertical bars represent the AUG initiation codons. Horizontal bars mark the length and direction of ORFs. Plots were derived by comparing the sequence against,

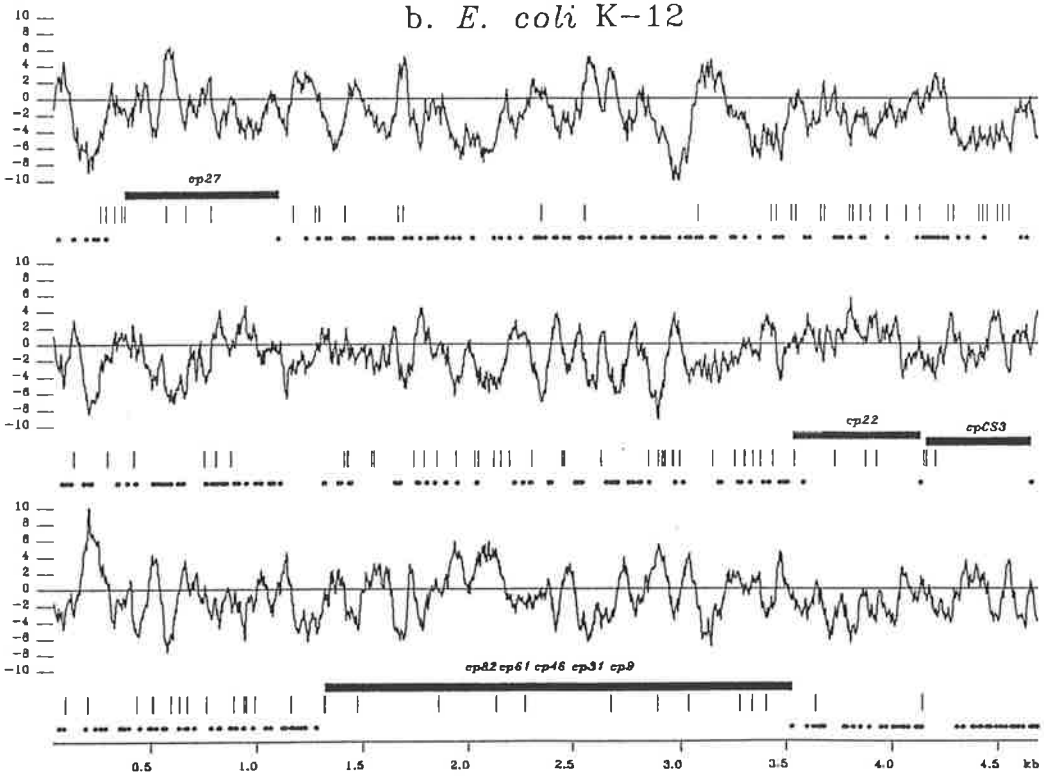
a - the ETEC codon frequency table (based on the genes, *fanD, fanE, faeD, faeE, fimD, fimE, papC, papD, papG* and *papH*)

b - the *E. coli* K-12 codon frequency table (Maruyama *et al.*, 1986)

a. ETEC



b. *E. coli* K–12



82150) of the polypeptide it encodes. The corresponding protein product is referred to as Cpnn (except the pilin subunit, which is referred to as CS3), and the potential promoter as P_{nn}. A summary of the ORFs is given in Table 3.1. These ORFs, as well as the potential promoters, a putative terminator and the relevant restriction sites are given schematically in Figure 3.6. All presumptive genes were on the same strand of the DNA. The largest ORF on the opposite strand, beginning with a UUG triplet, was only 65 amino acids long (MW 7057) and, therefore, this strand was considered to be non-coding. Of the eight ORFs found on the coding strand, four (*cp61*, *cp46*, *cp31* and *cp9*) were entirely contained in a fifth, and the largest, ORF (*cp82*) within the sequence. All five shared the same reading frame, using internal initiation codons and a common termination codon. The remaining three (*cp27*, *cp22* and *cpCS3*) were non-overlapping with any other.

The following sections describe the ORFs and properties of their deduced polypeptides:-

3.2.4.1 *cp27*

This is the first ORF on the sequence, which begins with an AUG (nt 378) and ends with a UGA (nt 1101), to give a polypeptide of MW 26893 (Figure 3.4a; Table 3.1). The string, GGAG, at nt 369 is appropriately located to function as an RBS. Although not ideal, the strings, TCAAAA (nt 312) and TATTAT (nt 336; Figure 3.4a; Table 3.2), show similarity to the -35 and -10 regions, respectively, of an *E. coli* σ^{70} promoter. The position of the last T residue in the -10 region is hereafter referred to as the sequence position of the putative promoters (e.g. nt 341 for P_{27}). Another potential promoter, with the -35 and -10 regions at nt 324 and 347, respectively, are also seen upstream to this ORF. The predicted size of this ORF (26.9 kDa) correlates well with the size of a protein band (27 kDa) seen in minicell labelling. Its location on the DNA also conformed with the prediction of Manning *et al.* (1985) for the 27 kDa polypeptide. Since it, most probably, represented the gene coding for the 27 kDa polypeptide, it was

Table 3.1

Summary of ORFs

ORF	RBS	Start	Stop	Mol. Wt.
cp27	GGAG-5-AUG	378	1101	26893
cp82	AAGA-6-AUG	1318	3523	82150
cp61	--	1858	3523	61171
cp46	AGGG-3-AUG	2266	3523	45861
cp31	GAAG-7-AUG	2668	3523	30687
cp9	GGAA-8-AUG	3274	3523	8668
cp22	--	3530	4130	21704
cpS3	GAGG-6-AUG	4154	4658	17475

Table 3.2

Summaries of potential promoters.

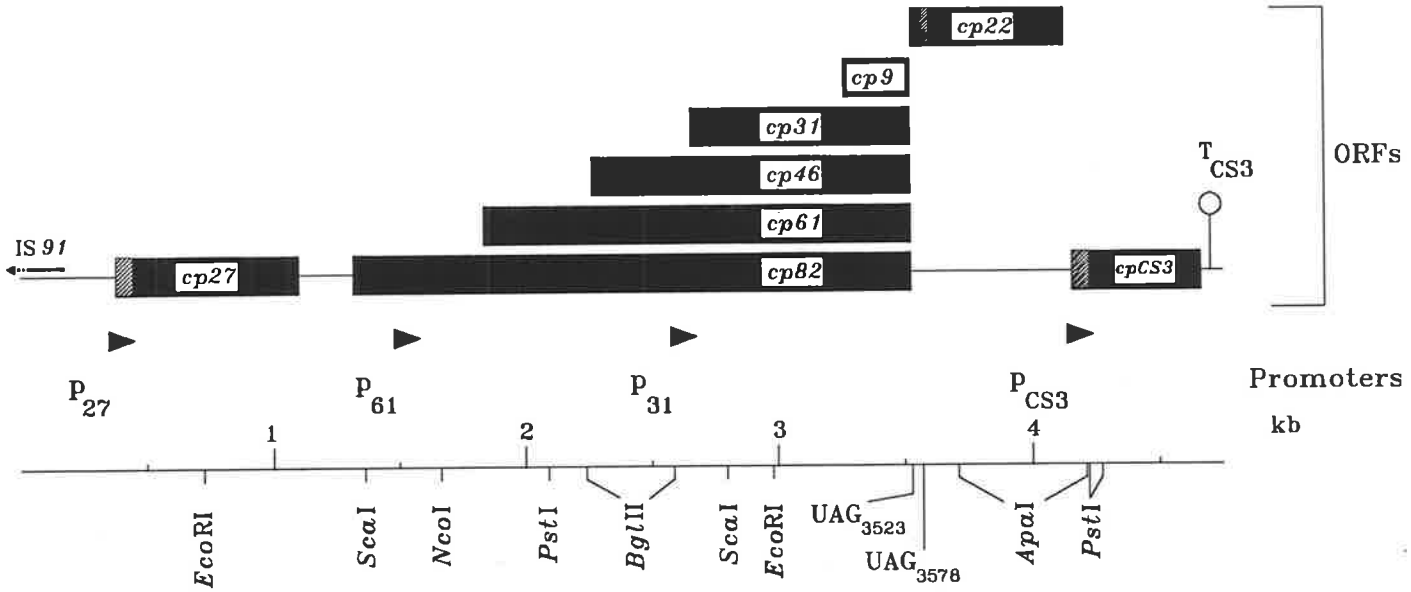
Seq. Pos.	Score	Designation*
341	-3.23	P_{27}#
352	1.06	
382	0.01	
631	0.05	
631	0.30	
822	1.56	
876	0.29	
1127	0.24	P_{82}
1154	0.19	
1479	0.75	
1797	-0.78	P_{61}#
1952	0.65	
2121	0.17	P_{46}
2433	0.00	
2574	0.35	P_{31}#
2702	0.15	
2718	0.79	
3252	-1.54	P_{9}
3482	0.21	
3483	1.15	P_{22}
3858	0.26	
4004	0.03	
4134	0.12	P_{CS3}#
4668	0.66	

Notes: * - The designation for promoters was arbitrary, based on their location on the DNA in relation to the ORFs.

- These promoters have been confirmed by primer extension or Northern blot analysis (*see* chapter 5).

Figure 3.6 Schematic diagram of ORFs and promoters.

Schematic representation of the potential ORFs. The DNA sequence is represented by a thin line, with the positions of relevant restriction sites marked. Solid boxes represent the ORFs, whereas shaded boxes indicate the putative signal peptides of proteins. Potential promoters (P_{27}, P_{61}, P_{31}, P_{CS3}) are marked by triangles. The potential terminator downstream to *cpCS3* (T_{CS3}) is also marked. The homology to IS*91* is also marked.


cp22
cp9
cp31
cp46
cp61
cp82
cp27
cpCS3
IS 91
T CS3
ORFs
P 27
P 61
P 31
P CS3
Promoters
kb
1
2
3
4
EcoRI
ScaI
NcoI
PstI
BglII
ScaI
EcoRI
UAG 3523
UAG 3578
ApaI
PstI

given the name, *cp27*. A leader peptide cleavage site between the residues 21 and 22 (Figure 3.4a) could be predicted by the -1,-3 rule of von Heijne (1983, 1986). Processing at this site would give a mature polypeptide of MW 24597, corresponding to the size of the shorter band seen in minicell translations. Codon usage within this ORF showed a bias towards the ETEC codons but not towards the *E. coli* K-12 codons (Figure 3.5).

3.2.4.2 *cp82*

The second ORF, termed *cp82* on the sequence, starts 215 nt downstream from the end of *cp27*. Translation initiating from any of the three inframe AUG codons at nt 1312, 1315 or 1318 would terminate at the UAG triplet at nt 3523 (hereafter referred to as UAG_{3523}) to give a polypeptide of MW 82150 (Table 3.1). The string, AAGA, at nt 1308 probably corresponds to its RBS (Figure 3.4a). Since an RBS is usually located 3 to 9 bases upstream from the initiation codon (Stormo *et al.*, 1982), the AUG triplet at nt 1318 appears to be the most likely start of translation. As in the case of *cp27*, the codon usage plot along this ORF shows a strong bias towards that for ETEC genes but poorly to that for *E. coli* K-12 genes (Figure 3.5). Although this ORF was the longest observed on the sequence in either direction, the polypeptide it encoded would be well short of the 97 kDa protein band observed in minicell translations. The sequence ladder in the vicinity of UAG_{3523} could be unambiguously read (*see* Figure 6.1, panel *b*), and the presence of stop codons in the other two reading frames discounted the possibility of a frameshift error that would make this ORF longer. Examination of the codon usage plot (Figure 3.5) revealed that the characteristic codon bias shifted the reading frame near the stop codon, suggesting that translation of this ORF terminated at this position. It appears, therefore, that an 82 kDa protein is migrating aberrantly on PAGE to give an apparent size of 97 kDa. Examination of the amino acid sequence of *cp82* did not reveal any unusual distribution of charged amino acids that might explain the abnormal mobility. In minicell translations, this band is closely followed by a shorter band of 94 kDa.

3.2.4.3 *cp22*

A faint band of approximately 22 kDa was consistently seen in minicell and *in vitro* translations during this study (*see* Figure 4.1, lane *b*), but had not been reported by Manning *et al.* (1985). A re-investigation of the translation pattern in pPM474, however, revealed the presence of this band. Two alternative start sites for its ORF are the AUG triplets at nt 2884 and 3530, respectively. Translation from the former terminates at the same UAG triplet that marks the end of *cp82*, to give a polypeptide of MW 22773. The initiation from the AUG at nt 3530 would only give a 2 kDa band. However, if the translation were to go past the UAG triplet nt 3578 (termed UAG_{3578}) a polypeptide of MW 21704 would result. Examination of the codon usage plot (Figure 3.5), which showed good coding potential across this UAG in the same reading frame, suggested that this area most likely encoded a protein. No initiation codon was present downstream to UAG_{3578}, to account for a 22 kDa protein. In addition to the coding potential across UAG_{3578}, the codon usage plot displayed the characteristic shift, from poor to good coding, near the AUG at 3530. In light of these indications, as well as based on other evidence given in Chapters 4 (section 4.2.1.6) and 6 (section 6.2.1), it was assumed that the translation of *cp22* initiated from the AUG at 3530, bypassed the UAG_{3578} and terminated at the UAA triplet at nt 4130.

3.2.4.4 *cpCS3*

The third open reading frame on the sequence starts with an AUG at nt 4154 and ends with a UAA at nt 4658 (Figure 3.4a and 3.6, Table 3.1). The string, GAGG, at nt 4144 appears to be the RBS, and the -35 and -10 regions, corresponding to a σ^{70} promoter, have been predicted at nt 4107 (TTAACG) and 4130 (TAAAGT), respectively (Figure 3.4a, Table 3.2). As in the case of *cp27* and *cp82*, this ORF also showed a bias towards the ETEC codon usage (Figure 3.5). Judging from its size (17.5 kDa) and relative location to the physical map of pPM474 (Manning *et al.*, 1985), it appeared to code for the structural subunit of CS3 and hence was designated *cpCS3*.

While this work was in progress, Boylan *et al.* (1988b) reported the sequence of the DNA encoding the CS3 gene from the ETEC strain, E90a. The DNA and deduced amino acid sequences reported by these authors are identical to those presented here except for one conservative amino acid change, within *cpCS3* at nt 4400 (Ser instead of Asn) and two base pair differences (G instead of A, in both cases) at nt 4089 (upstream to *cpCS3*) and nt 4401 (Figure 3.4a). However, the restriction maps of the *Hin*dIII fragment, encoding all genes necessary for the biosynthesis of CS3 pili, do differ from each other (see Figure 2 in Boylan *et al.*, 1987).

3.2.5 Overlapping translation is needed for other ORFs

Almost the entire DNA in one direction, from nt 378 to nt 4660, has been covered by the four ORFs (Figure 3.6) described above. Other reading frames in the same or opposite directions were devoid of any ORF of significant length. As the largest ORF on the opposite strand, beginning with a UUG triplet, was only 65 amino acids long (MW of 7057), this strand was not considered to be coding for any of the protein bands seen in minicell labelling. Manning *et al.* (1985) have considered two alternatives for the production of the 58 and 46 kDa polypeptides:-

(1) specific cleavage of one or more of the larger polypeptides and

(2) re-initiation of translation within the coding region of a larger polypeptide.

Assuming re-initiation of translations, rather than post-translational processing, to produce the remaining four polypeptides, the DNA sequence was searched for potential initiation codons within the four identified ORFs (*i.e. cp27*, *cp82*, *cp22* and *cpCS3*). Although the 9 kDa band was small enough to be contained within *cp27*, *cp22* or *cpCS3*, the absence of suitable initiation codons within these ORFs ruled out this possibility. However, *cp82* showed several potential initiation codons, and these were critically examined in the search for the ORFs for the 61, 46, 31 and 9 kDa proteins. The following sections describe these ORFs.

3.2.5.1 *cp61*

Initiation of translation from the GUG at nt 1837 or the AUG at nt 1858 (Figure 3.4a and 3.6) would give rise to a polypeptide of MW 61971 or 61171, respectively (Table 3.1). The AUG triplet was taken as the likely initiation codon for the 61 kDa protein which, in turn, was assumed to be the 58 kDa band reported by Manning *et al.* (1985). No suitable RBS was found before this codon.

3.2.5.2 *cp46*

Since the difference in size between the 46 kDa and the 61 kDa bands was too large to consider them to be the precursor and mature forms of the same gene product, a potential translation initiation site for the 46 kDa polypeptide was sought. Translation from the AUG codon at nt 2266 would produce a polypeptide of MW 45861. The string, AGGG, at nt 2259 appears to be an RBS for its translation (Figure 3.4a; Table 3.1). It should be noted that the 46 kDa band was faint in minicell labelling, possibly reflecting a weak RBS.

3.2.5.3 *cp31* and *cp9*

The last two ORFs to be examined, like *cp22*, corresponded to protein products not reported by Manning *et al.* (1985). The intensities of their protein bands in minicell translations were higher than those of *cp61* and *cp46*, suggesting the presence of either a stronger RBS and/or a promoter to reinitiate transcription. Alternatively, their mRNAs may have more stability.

cp31

Two alternative start sites to produce a 31 kDa polypeptide were found at nt 291, upstream to the start site of *cp27*, and at nt 2668 within the reading frame of *cp82*. Though both initiation codons were preceded by the string, GAAG, that could be considered as an RBS, examination of the codon usage plot (Figure 3.5) showed poor coding potential in the vicinity of the AUG at nt 291. The ORF starting at 2668 is within

cp82, and therefore shared a better coding region with the latter (Figure 3.5). This ORF is also preceded by a putative promoter, with the strings GTGGAA (nt 2546) and GAGAAT (nt 2569) as the potential -35 and -10 regions (Figure 3.4a; Table 3.2). The string, GAAG, at nt 2657 appeared to be its RBS (Figure 3.4a; Table 3.1). The AUG triplet at nt 2668 was therefore taken as the most likely initiation codon for the 31 kDa polypeptide.

cp9

The smallest polypeptide seen in minicell and *in vitro* translations had a size of 9 kDa and was seldom observed, since it often ran off the gel during electrophoresis. The AUG at nt 3274 was taken as the candidate for its initiation codon. The string, GGAA (nt 3262), preceding this triplet would form a likely RBS. Although a good -10 region (string TATAAT at nt 3247) was found before this ORF, the -35 region (string CTGGTG at nt 3224) showed only poor homology to the σ^{70} consensus (Figure 3.4a; Table 3.1 and 3.2).

3.2.6 Other Features of the sequence

Searches against nucleotide databases revealed several interesting homologies. Other sequence analyses revealed potential signals, like promoters and protein binding regions.

3.2.6.1 Sequence homologies with other DNAs

The entire 4747 bp DNA sequence was compared against the GenBank and EMBL nucleotide databases, using the FASTA program (Pearson and Lipman, 1988), to detect probable homologies to other pilus biosynthesis genes. An arbitrary value of 100 for the optimised score of the program's output was used as the cut-off point when selecting the homology alignments for further study. Only those homologies which are statistically or functionally significant are described in the following sections.

3.2.6.1a IS*91* element

Good sequence homology was observed with the insertion element IS*91* between nt 68 and 167 (Figure 3.7a). Since only about 100 bases from each end of the IS*91* element (about 1.85 kb) have been sequenced (Diaz-Aroca *et al.*, 1987), it could not be verified whether nt 1 to 67 of the ETEC sequence also matched it. From the observed sequence homology, it is apparent that at least a part of IS*91*, or a closely related IS element, is present on the plasmid, in such an orientation that its left end is within the DNA cloned in pPM484. If the entire IS*91* element is present in the wild type plasmid, then, its right end must be approximately 1.85 kb to the left of the *Hin*dIII (nt 1) that marks the left end of the insert DNA in pPM484.

3.2.6.1b *cfaA*

Strong homology was observed between the promoter regions of *cp27* and the *cfaA* gene (Hamers *et al.*, 1989) in the *E. coli* CFA/I fimbrial operon (Figure 3.7b). A better consensus for the -10 region of the promoter, than the one marked by Hamers *et al.* (1989), was identified in front of the *cfaA* gene sequence, and this exactly overlapped with P_{27}. Hamers *et al.* (1989) have also observed a 172 bp fragment of the IS2 element (about 1.1 kb; Ghosal *et al.*, 1979) located at 128 bp upstream to the *cfaA* promoter. The end of IS*91* element in pPM484 lies at 174 bp upstream to P_{27} and, therefore, is approximately in the same relative location as the IS2 element is to *cfaA*. There is no sequence similarity between IS*91* and IS2 elements.

3.2.6.1c *E. coli* adhesin gene, F17-G

Similarity to a region 600 bp downstream of the *E. coli* adhesin gene (F17-G; Lintermans *et al.*, 1991) was found between 76 and 135 base pairs upstream to the *cp27* promoter. This homology breaks down near the junction between the IS*91* element and pPM484 sequence. No homology was seen between the sequences around the two promoters (Figure 3.7c).

Figure 3.7 DNA sequence homologies

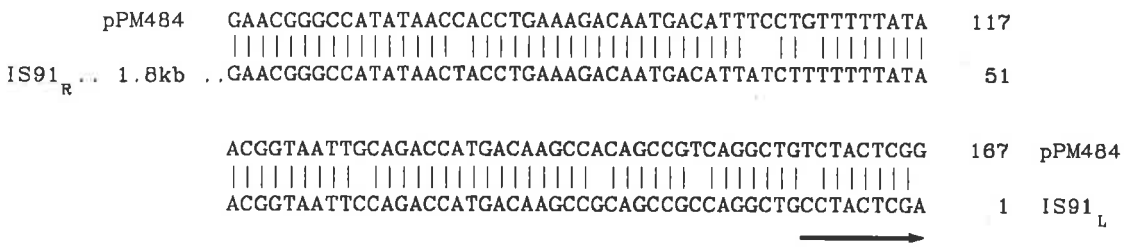
Homologies detected by comparison against nucleotide databases.

a - Against the IS*91* element. The complementary strand of the left end of IS*91* is shown here. Arrow indicates the terminal inverted repeat at the ends of IS*91*.

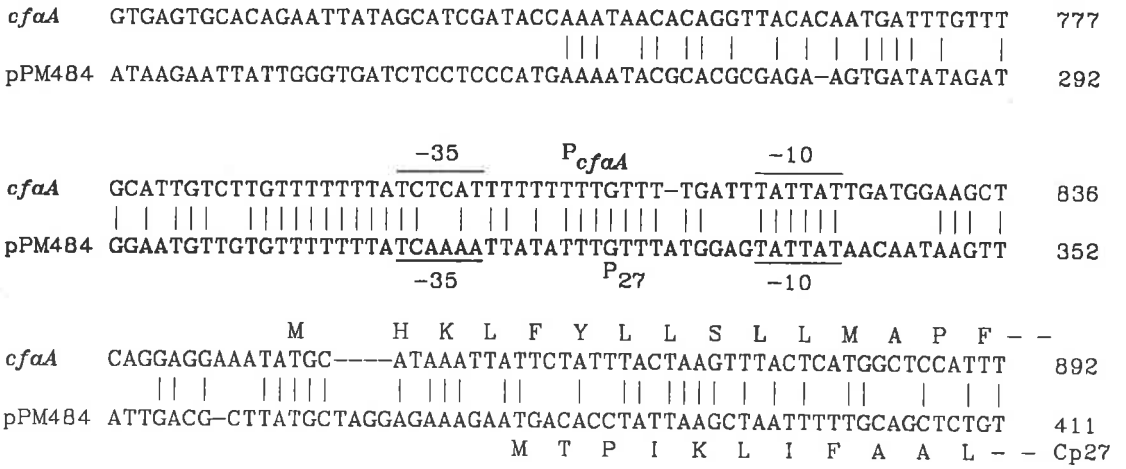
b - Against the promoter region of the *cfaA* gene in the CFA/I operon. The deduced -35 and -10 regions of the two promoters, P_{27} and P_{cfaA}, are overlined. Partial amino acid sequences of the two genes are also shown.

c - Against the *E. coli* adhesin gene, F17-G.

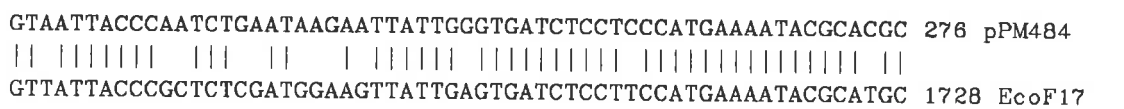
a


pPM484 GAACGGGCCATATAACCACCTGAAAGACAATGACATTTCCTGTTTTTATA 117
IS91R .. 1.8kb ..GAACGGGCCATATAACTACCTGAAAGACAATGACATTATCTTTTTTTATA 51
ACGGTAATTGCAGACCATGACAAGCCACAGCCGTCAGGCTGTCTACTCGG 167 pPM484
ACGGTAATTCCAGACCATGACAAGCCGCAGCCGCCAGGCTGCCTACTCGA 1 IS91L

b


cfaA GTGAGTGCACAGAATTATAGCATCGATACCAAATAACACAGGTTACACAATGATTTGTTT 777
pPM484 ATAAGAATTATTGGGTGATCTCCTCCCATGAAAATACGCACGCGAGA–AGTGATATAGAT 292
−35 PcfaA −10
cfaA GCATTGTCTTGTTTTTTTATCTCATTTTTTTTTGTTT–TGATTTATTATTGATGGAAGCT 836
pPM484 GGAATGTTGTGTTTTTTTATCAAAATTATATTTGTTTATGGAGTATTATAACAATAAGTT 352
−35 P27 −10
M H K L F Y L L S L L M A P F – –
cfaA CAGGAGGAAATATGC-----ATAAATTATTCTATTTACTAAGTTTACTCATGGCTCCATTT 892
pPM484 ATTGACG–CTTATGCTAGGAGAAAGAATGACACCTATTAAGCTAATTTTTGCAGCTCTGT 411
M T P I K L I F A A L – – Cp27

c


GTAATTACCCAATCTGAATAAGAATTATTGGGTGATCTCCTCCCATGAAAATACGCACGC 276 pPM484
GTTATTACCCGCTCTCGATGGAAGTTATTGAGTGATCTCCTTCCATGAAAATACGCATGC 1728 EcoF17

Besides the above two homologies, which could have some functional significance, several others were noted. Most of these occurred in A+T-rich regions and were probably spurious.

3.2.6.2 Promoters

A computer search, using a score matrix method (Kalionis *et al.*, 1986), identified an unusually high number of potential promoters on the sequence (Table 3.2). The most likely candidates are marked on the sequence at nt 341, 1797, 2574 and 4135. These are designated P_{27}, P_{61}, P_{31} and P_{CS3}, respectively (Figure 3.4a and 3.6).

3.2.6.3 Terminators

Weak stem-loop structures with runs of 3 to 5 Us were found at nt 387, 1775 and 1967. Judging from the location on the sequence, and the relatively positive free energy of the stem-loops, these did not appear to be candidates for Rho-independent terminators. A potential stem-loop structure, followed by a U-rich region, but not an uninterrupted run of Us, appears at nt 4671 with a free energy of -10.8 kcal/mol (Tinoco *et al.*, 1973) This appears to be a candidate for the terminator of CS3 mRNA (Figure 3.4a and 3.8).

3.2.6.4 Direct repeats

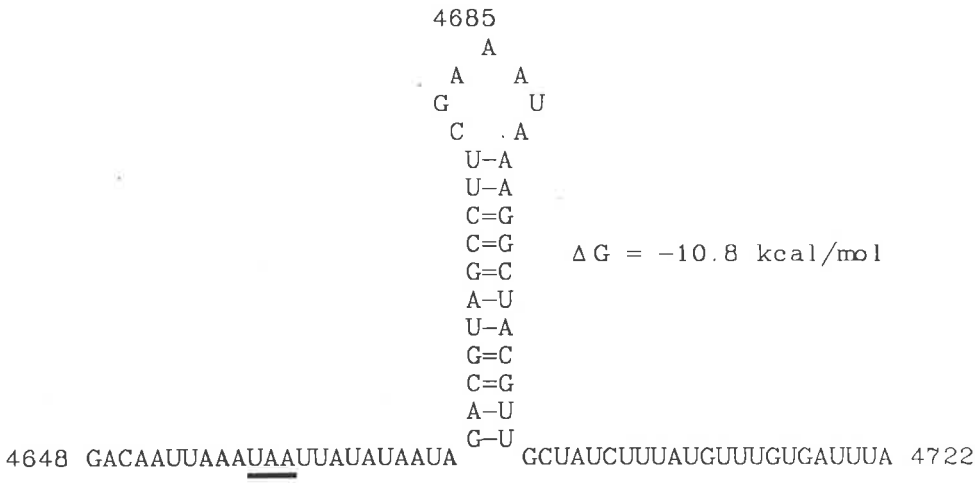
The minimum length for a statistically significant repeat on the 4747 bp DNA of pPM484 was 12 bases. The longest repeat (TTTAATGGAAAGGC) was 14 bp in length. These were found at nt 948, within the coding region of *cp27*, and at nt 2620, immediately upstream to *cp31* (Figure 3.4a). This string is not known to have any functional significance.

3.2.6.5 Protein binding sites

The string AAAAAATTGATT, at nt 1218, shows 8/9 homology with the consensus sequence for IHF binding (AANNNRTTGATR; Nash,1981; Friedman, 1988).

Figure 3.8 Transcriptional terminator in the CS3 operon.

Potential stem-loop structure downstream of *cpCS3*. The free energy of the structure was calculated by the method of Tinoco *et al.* (1973). The termination codon of *cpCS3* is underlined.


4685
A
A A
G U
C A
U–A
U–A
C=G
C=G
G=C
A–U
U–A
G=C
C=G
A–U
G–U
ΔG = –10.8 kcal/mol
4648 GACAAUUAAAUAAUUAUAUAAUA GCUAUCUUUAUGUUUGUGAUUUA 4722

This site precedes the translation initiation site of *cp82* (Figure 3.4a). No other protein binding site was found on the sequence.

3.2.6.6 G+C content and Restriction Enzyme Sites

The DNA sequence of pPM484 displayed a relatively low G+C content (36%), and in this regard deviated significantly from *E. coli* K-12, where the G+C and A+T contents are approximately equal (50%; Caron *et al.*, 1989). In order to verify whether the low G+C content was a characteristic of pilus biosynthesis genes, the nucleotide sequences of genes from other pilus systems were extracted from the GenBank database and examined. The percentage of G+C and the GenBank accession numbers for the respective sequences are given in brackets. A low G+C content was seen in *cfaA* (33%; M55661), *fanD* (34%; X13560), *fanE* (34%; X56001) *papE* (40%; M13239) and *papF* (40%; M13239) genes. In contrast, the G+C content is higher in *faeD* gene (58%; X03675). The G+C/A+T contents were approximately equal in the *faeE* (X56003), *fimG* (X05672), *fimH* (X05672), *papD* (M30806), *papF* (M13239) and *papH* (M16202) genes.

Overall, the frequencies of restriction sites in the DNA reflected the low G+C content; *i.e.* sites containing only G and C nucleotides were fewer than those containing all four nucleotides which, in turn, were fewer than those containing A and T alone. Sites of all known enzymes with 4-base recognition sequences were present on the fragment, as would be expected on a 4.7 kb fragment, the lowest number being one for *Fnu*DII (CGCG) and the highest being 47 for *Mse*I (TTAA).

The relevant restriction sites that were used in the mutagenesis of ORFs, as well as those sites mapped by Manning *et al.* (1985), are listed in Table 3.3 and underlined on the sequence in Figure 3.4a. Two *Bgl*II sites, instead of the lone site reported by Manning *et al.* (1985), were seen on the sequence. Instead of the two reported *Pst*I sites, there were three on the sequence. The short distances between the two *Bgl*II sites (348 bp), and two of the three *Pst*I sites (51 bp) would have been responsible for these being

Table 3.3

Summary of Relevant Restriction sites

Restriction Site	Seq. Position(s)		
*Apa*I	3706	4226	
*Bgl*II	2239	2587	
*Eco*RI	724	2978	
*Nco*I	1663		
*Pst*I	2087	4218	4269
*Sca*I	1358	2788	

classified as single sites in earlier studies. One of these two *Pst*I sites occurred in the area encoding the signal peptide cleavage site of CS3 polypeptide and the other at 51 bases downstream from it. In future, these sites could become useful in either a deletion analysis of CS3 subunit or the N-terminal fusion of a foreign antigen. A lone *Cla*I site at nt 3084 overlaps a *dam* methylase recognition site and, due to methylation, was uncleavable when pPM484 DNA was prepared from a *dam*$^+$ *E. coli* K-12 strain.

3.3 Discussion

This chapter firstly investigated the ability to mediate CS3 biosynthesis by the plasmid, pPM484, which does not contain the additional 0.7 kb *Hin*dIII fragment found in the plasmid, pPM474, originally used by Manning *et al.* (1985). Minicell labelling was used to confirm that pPM484 encoded all of the CS3 biosynthesis-related polypeptides reported to be present in pPM474. Three additional bands with sizes of 31, 22 and 9 kDa were discovered during this study. The probable reasons for the failure to detect these bands earlier studies are,

(1) masking by the Ap^R gene product (31 kDa),

(2) poor expression (22 kDa), and

(3) small size (9 kDa).

The entire DNA sequence of the insert was determined in order to facilitate the identification of the CS3 biosynthesis genes. The sequence was unremarkable in terms of structures and repeats, but showed an unusually extensive overlapping of translation. Some overlap is common in prokaryotic transcription and translation, which was anticipated to be imperative in the case of the CS3 gene cluster (Manning *et al.*, 1985). However, the extent of overlap at the level of translation, and the possibility also at transcription, was unusual. Five presumptive genes overlapped with each other. Four of these were completely contained within the largest gene, termed *cp82*.

Previous evidence for the existence of several genes came from minicell translations as well as physical mapping of the DNA by transposon mutagenesis and

*Bal*31 deletion (Manning *et al.*, 1985). Interpretation of results from these experiments was complicated by the apparent overlapping of translation, and alternative explanations were possible. For example, a band seen in minicell translation could also have formed by post-translational processing or by degradation of the product before PAGE analysis. Because of the genetic organization, deletion derivatives and transposon mutants gave polar effects which did not permit an accurate picture of the number and location of translational units. In this chapter, an attempt has been made to locate the genes by DNA sequence analysis. It has been assumed that the major bands seen in minicell translations are produced by independent initiations of translation rather than post-translational processing. Although codon usage patterns generally give an accurate assessment of the coding regions of genes, it could not be used in the identification of *cp61*, *cp46*, *cp31*, *cp22* and *cp9*, all of which were contained within the reading frame of *cp82*. These genes were therefore assigned by taking into account the size of the polypeptides that could be formed by translational initiations from the putative initiation codons. The presence of potential RBS and promoter sequences were also considered in these assignments. A complication in the latter was that homologies to these signals were not strong. Of relevance is that several known genes in *E. coli* K-12 are found to lack a distinct RBS or promoter (Gold *et al.*, 1981; Hershey, 1987; Gold and Stormo, 1987).

The DNA sequence revealed, as predicted, that the 27 kDa protein was not derived from any other polypeptide and had independent transcription and translation signals (Manning *et al.*, 1985). Similarly, the CS3 pilin subunit also had its own putative promoter and translation initiation sites. The 22 kDa protein appeared to be translated independently, but was probably encoded on the same mRNA that encoded the upstream ORFs. An in-frame UAG triplet is found in its presumptive ORF. Evidence to believe that AUG at nt 3530 is in fact used as initiation site of translation of *cp22*, and UAG_{3578} is either suppressed or bypassed during translation, comes from the following observations. Firstly, this area on the sequence had been verified thoroughly to rule out

any possible sequencing error. Secondly, the codon usage plot across this codon indicated good coding potential, with no potential initiation codon downstream to UAG_{3578} to give rise to a polypeptide of approximately 22 kDa. Thirdly, cell surface pili are absent in an Su^- strain (*see* Chapter 6), which indicates that suppression of an amber codon, presumably the one in *cp22*, plays a part in pilus biogenesis.

The DNA sequence was devoid of extensive direct or inverted repeats. An unusually large number of potential σ^{70} promoters were identified on both strands of the DNA. The high A+T content of the DNA (64%) could have been responsible for detecting a large number of spurious promoters by a computer algorithm, which used a score matrix containing high frequencies of A and T residues at several positions (overall A+T content in the matrix = 62.3%). Owing to the functional significance of their locations, the promoters found in front of *cp27* and *cpCS3* are probably genuine. However, the probability score for these two promoters was lower than several others found on either strand of the DNA. Analysis of transcription from these computer-predicted promoters, by Northern blots or *in vivo* activity assays, is the best way to confirm or rule out their existence.

The presence of a putative IHF binding site in front of *cp82* suggests a possible role of this host protein in regulating its translation. Examining the production of CS3 pilus in a strain defective in the *ihf* gene may resolve this.

Proving the overlapping translations in the CS3 biosynthesis gene cluster is imperative, and forms a major part of the studies reported in this thesis. Availability of the nucleotide sequence will greatly facilitate the introduction of specific mutations into the DNA. End-filling single restriction sites to create frameshift mutations in their coding regions may help to identify the possible initiation sites of the polypeptides. The spatial organization of the non-overlapping genes suggests that they form a single operon. Transcription of the downstream genes could therefore be dependent upon the

translation of upstream genes. Possible polarity of the upstream frameshift mutations must be taken into consideration when analyzing their effects.

Frameshift mutants may also be assessed to verify the role of the various gene products in pilus biosynthesis. A difficulty in interpreting these results is that all mutations that abolish the downstream genes (*cp61*, *cp46*, *cp31* and *cp9*) may also result in a mutation of *cp82* and, if the latter is essential for CS3 fimbrial synthesis, will result in the loss of CS3 expression. These questions are addressed in the next chapter, which deals with the experimental confirmation of the proposed initiation codons, computer analyses of the proteins, verification of the involvement of all proteins in pilus biogenesis, and the determination of their potential roles.

CHAPTER 4

Identification, Analyses And Roles Of The CS3 Pilus Biosynthesis Proteins

CHAPTER 4

Identification, Analyses And Roles Of The CS3 Pilus Biosynthesis Proteins.

4.1 Introduction

Predictions, based on the DNA sequence data for the region encoding CS3 biosynthesis, suggested an overlapping gene arrangement to account for the proteins detected by minicell labelling. Four ORFs were fully contained within a larger ORF, *cp82*, and used a common termination codon, but different start codons. These ORFs were arbitrarily defined, based on the sizes of polypeptides and the presence of likely ribosome binding sites (RBS) preceding their initiation codons. To confirm these predictions it was necessary to introduce alterations into the DNA, to verify whether they produced the expected effect on translations. Two approaches considered were progressive deletions from ends of the DNA, and site-specific mutagenesis of the coding regions. The former is useful when the genes are independently transcribed and translated. Since it was probable that most of the genes in pPM484 were transcribed from a single upstream promoter, deleting the DNA from the 5' end would affect all genes if this promoter was deleted or inactivated. Similarly, a common termination codon was being used by the overlapping genes and, therefore, deleting the DNA from the 3' end could also lead to erroneous conclusions. For these reasons, it was decided that site-specific mutations in the coding regions were more appropriate in determining the starts and ends of genes, as well as investigating the roles of their products.

This chapter analyses the eight potential genes encoded within pPM484, in terms of their location in the DNA and the involvement of their products in pilus biosynthesis. The initiation and termination sites were mapped by introducing site-specific frame-shifts into the coding regions, followed by minicell labelling. The effect of these mutations on pilus production was investigated to verify the involvement of these proteins in its biogenesis. Finally, an attempt has been made to assign possible roles to these proteins by

similarity to other known proteins, and by investigating their intracellular locations using cell fractionation studies. The sequence positions mentioned in this chapter correspond to those in Figure 3.4a of Chapter 3.

4.2 Results

4.2.1 Correlation of protein bands to ORFs.

Due to their overlapping arrangement, a number of potential initiation sites had to be considered for each gene. Absolute proof for the translation initiation sites could only be obtained by mutagenizing these predicted initiation codons. Indirect evidence, however, could be obtained by site-directed frameshift mutations that resulted in a truncation or loss of the polypeptides. Either an insertion of 4 bases, in the case of *Eco*RI and *Nco*I (restriction enzymes that produce 5' overhang), or a deletion of 4 bases, in the case of *Pst*I (enzyme that produces 3' overhang), was created by digesting with the respective enzyme, treating with Klenow (in the presence of all four dNTPs) and re-ligating with T4 DNA ligase. The following sections describe, firstly, the correlation of a minicell-identified protein band to an ORF and, then, mutational analyses to confirm its start of translation.

4.2.1.1 Cp27

Two protein bands, migrating slightly ahead of the mature form of β-lactamase, are the candidates for the *cp27* products (Figure 4.1, lane *c*). Only the larger of these bands is found in *in vitro* translation of pPM484 DNA (Figure 4.1, lane *b*), or the *in vivo* (minicell) translation in the presence of 8% ethanol (Figure 4.2, lane *c*). Since leader peptide processing does not occur in these methods of translation, absence of the shorter band suggests that it represents the mature form of Cp27 protein.

One of the three *Eco*RI sites in pPM484 is within *cp27* at nt 724. End-filling this site would result in a frameshift and produce truncated proteins with sizes of 14.3 and 12.0 kDa, respectively, for the precursor and mature polypeptides. The desired mutation

Figure 4.1 Comparison of the *in vitro* and minicell translation of pPM484

Translation products labelled with ^{35}S-methionine were run in SDS on 15% PAGE, flourographed and exposed overnight at -80°C.

Lanes:

a - *In vitro* translation of pBR322 DNA

b - *In vitro* translation of pPM484 DNA

c - *In vivo* (minicell) translation of pPM484

d - *In vivo* (minicell) translation of pBR322.

Potential gene products are marked. Precursor forms of the proteins are indicated by bi-directional arrows whereas mature products in the minicell's lane are indicated by short arrows. The product of *cp31* appears to co-migrate with *pre-bla* in the *in vitro* translation, but slightly below it in the *in vivo* system. Molecular weights, as determined from the mobility of pre-stained protein MW standards are marked.

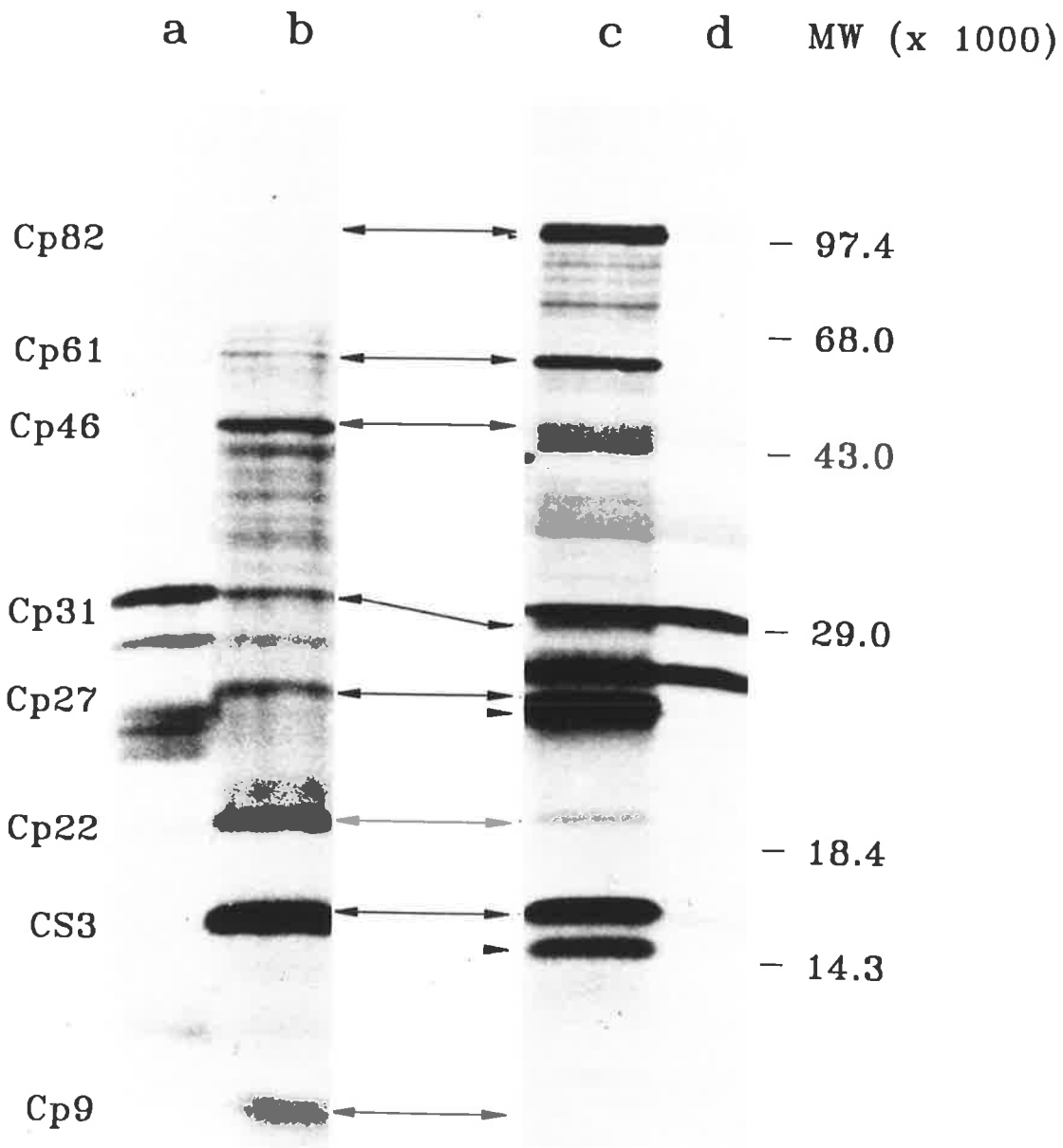

a
b
c
d
MW (x 1000)
Cp82
Cp61
Cp46
Cp31
Cp27
Cp22
CS3
Cp9
97.4
68.0
43.0
29.0
18.4
14.3

Figure 4.2 Minicell translation in the presence of 8% ethanol

Minicells containing either pBR322 or pPM484 were labelled in the presence or absence of 8% ethanol. The bands for Cp27, Cp22, CS3, Cp9 and β-lactamase (bla) are marked. The presence of non-specific bands in all tracks made it dificult to assess whether the observed bands for Cp82, Cp61, Cp46 and Cp31 were real and, hence, their expected positions are given in brackets.

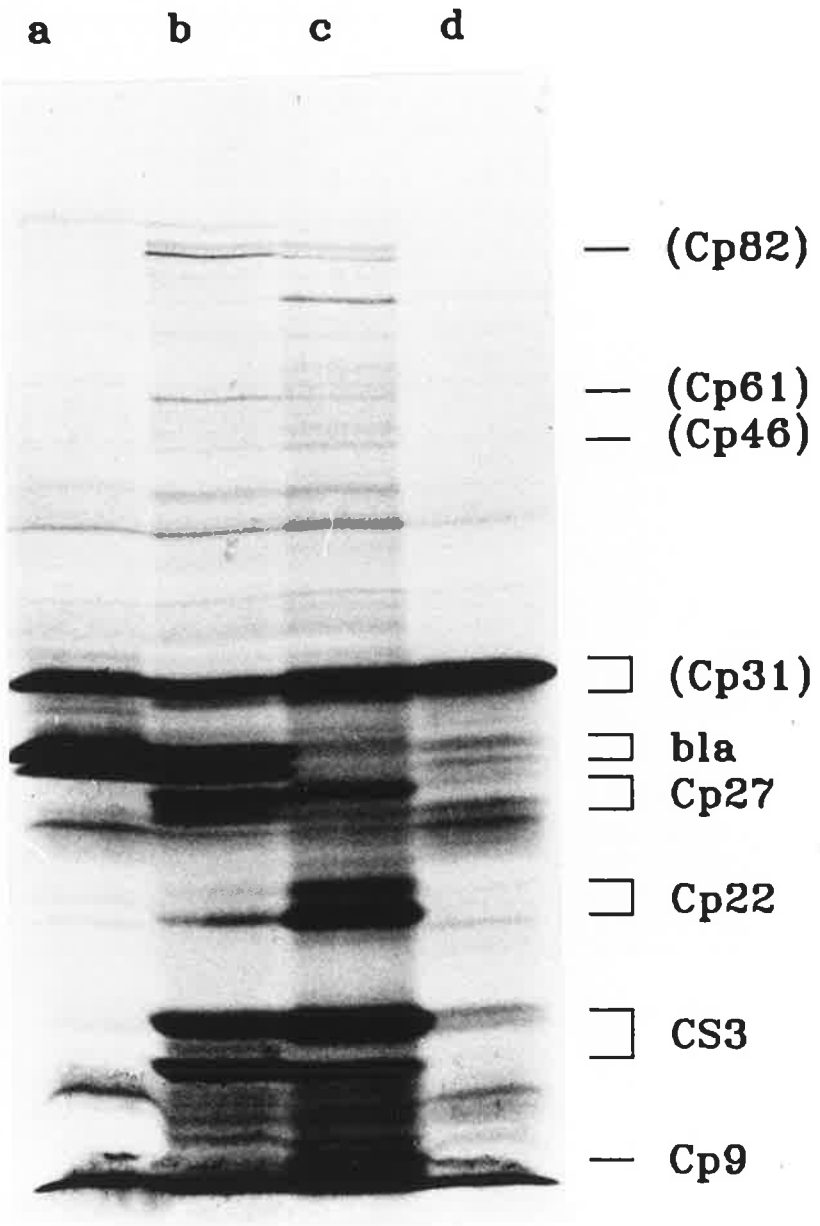
Lanes:

a - pBR322 in the absence of ethanol.

b - pPM484 in the absence of ethanol.

c - pPM484 in the presence of 8% ethanol.

d - pBR322 in the presence of 8% ethanol.


a
b
c
d
(Cp82)
(Cp61)
(Cp46)
(Cp31)
bla
Cp27
Cp22
CS3
Cp9

(hereafter referred to as the *Eco*RI mutation) was achieved by partially digesting the plasmid DNA with *Eco*RI, end-filling with Klenow and re-ligating. Plasmids from several colonies were screened to identify the one that had lost the *Eco*RI site at nt 724 but retained the other two sites. Translation *in vivo* revealed that the two bands originally assigned to Cp27 disappeared, coinciding with the appearance of two faster migrating bands (Figure 4.3A, lane *b*). The truncated precursor polypeptide (predicted size of 14.3 kDa) migrated more slowly than the mature form of CS3 subunit (15.8 kDa), giving the former an apparent size of 17 kDa. This raised the possibility that the actual initiation site of Cp27 was upstream to the previously predicted site at nt 378. The following observations, however, ruled out this possibility:-

Firstly, even though the truncated product would be approximately 17 kDa, if translation were to initiate from an alternative AUG at nt 291, the full length Cp27 precursor protein (predicted size of 30.4 kDa) in this case should migrate more slowly than the mature β-lactamase (29.7 kDa). In reality, however, it migrated ahead of the latter.

Secondly, the predicted signal peptide cleavage site after the aa residue 21 (Figure 3.4a), if translation began from the AUG at nt 378, made it the favoured candidate. No signal peptide cleavage site could be predicted at a similar distance from the AUG at nt 291.

Thirdly, the presence of a potential promoter at nt 341 made the AUG at nt 378 (*see* Chapter 5, Section 5.2.3) the most likely candidate, as no potential promoter was found upstream of the alternative AUG triplet.

Fourthly, a characteristic shift of the codon usage plot (Figure 3.5) from poor to good coding potential occurred near the AUG at nt 378.

A possible explanation for the disparity between the predicted and observed sizes of the truncated products of Cp27 was that translation after the frame-shift mutagenesis of *cp27* might be terminating further downstream than as suggested by the sequence. A

Figure 4.3A Minicell translation of the mutant derivatives of pPM484

Minicells prepared from DS410, transformed with pPM484 and its mutant derivatives were labelled with ^{35}S- Met *in vivo* and separated in SDS on 15% PAGE.

Lanes:

a - pPM484

b - The *Eco*RI mutant (pPM1353)

c - The *Bgl*II mutant (pPM1352)

d - pBR322

e - pPM484

f - The *Nco*I mutant (pPM1351)

g,h - pBR322

i - pPM484

j - The *Bgl*II mutant (pPM1352)

k - The *Pst* I mutant (pPM788)

l - pPM484

m - pBR322

Accumulation of Cp22 as a result of the *Bgl* II mutation is visible in lane *j* whereas Cp46 is accumulated as a result of the *Pst* I mutation (lane *k*). The band for Cp31 is seen above that of pre-β-lactamase in lanes *a* to *c* whereas in lanes *k* and *l* its migrates faster. To indicate this fact the observed band, as well as its expected position, in lanes *k* and *l* are marked. Molecular weights (kDa) are marked between the lanes *d* and *e*.

A

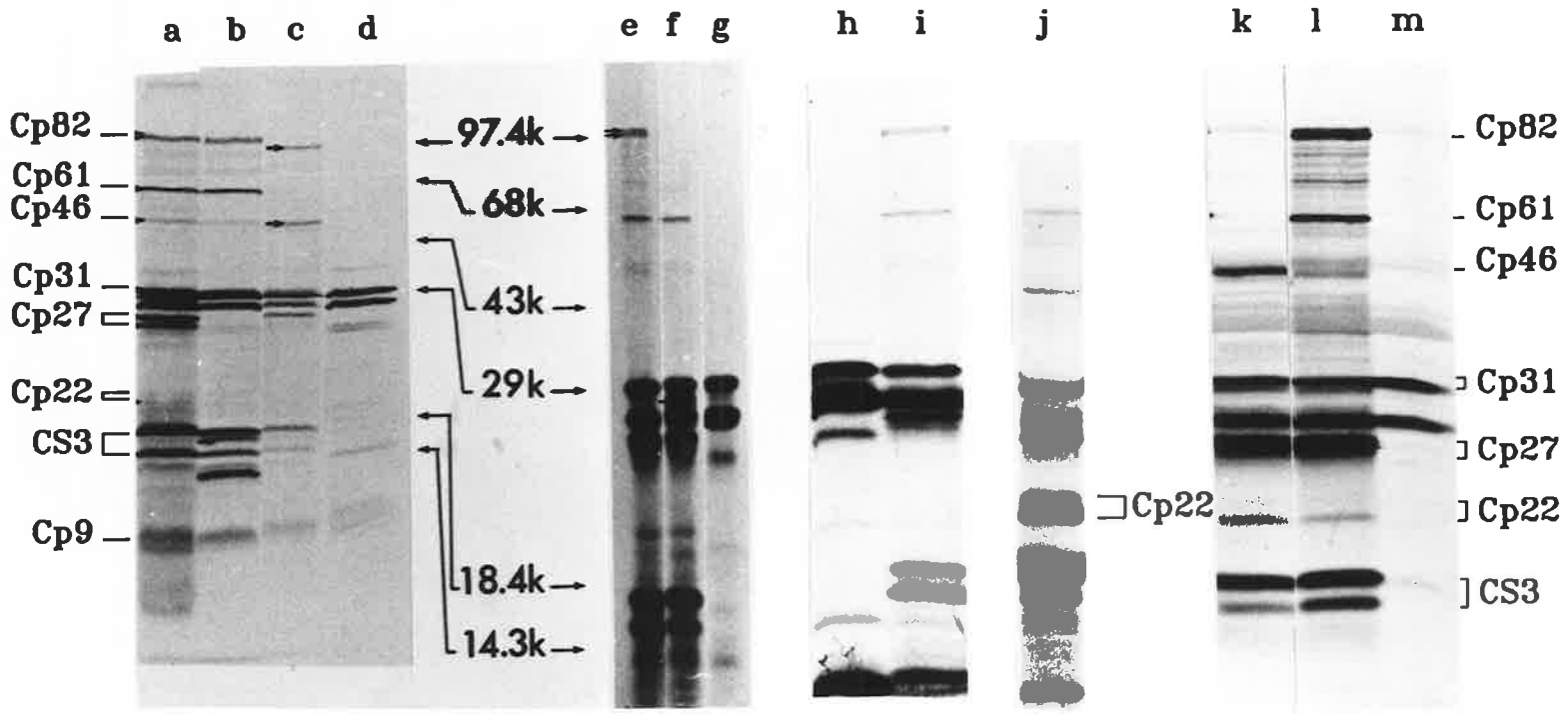

a
b
c
d
e
f
g
h
i
j
k
l
m
Cp82
Cp61
Cp46
Cp31
Cp27
Cp22
CS3
Cp9
97.4k
68k
43k
29k
18.4k
14.3k
Cp22
Cp82
Cp61
Cp46
Cp31
Cp27
Cp22
CS3

Figure 4.3B Minicell translation of the *EcoRI* and the *Bgl*II deletion mutants

This figure is an expanded view of the first four lanes in figure 4.3A to show the band assigned to Cp31. The truncated product of Cp61, associated with the *Bgl*II deletion is marked as Δ Cp61 in lane c. Also marked are the Cp protein bands and the two forms of β-lactamase (pre-bla and bla).

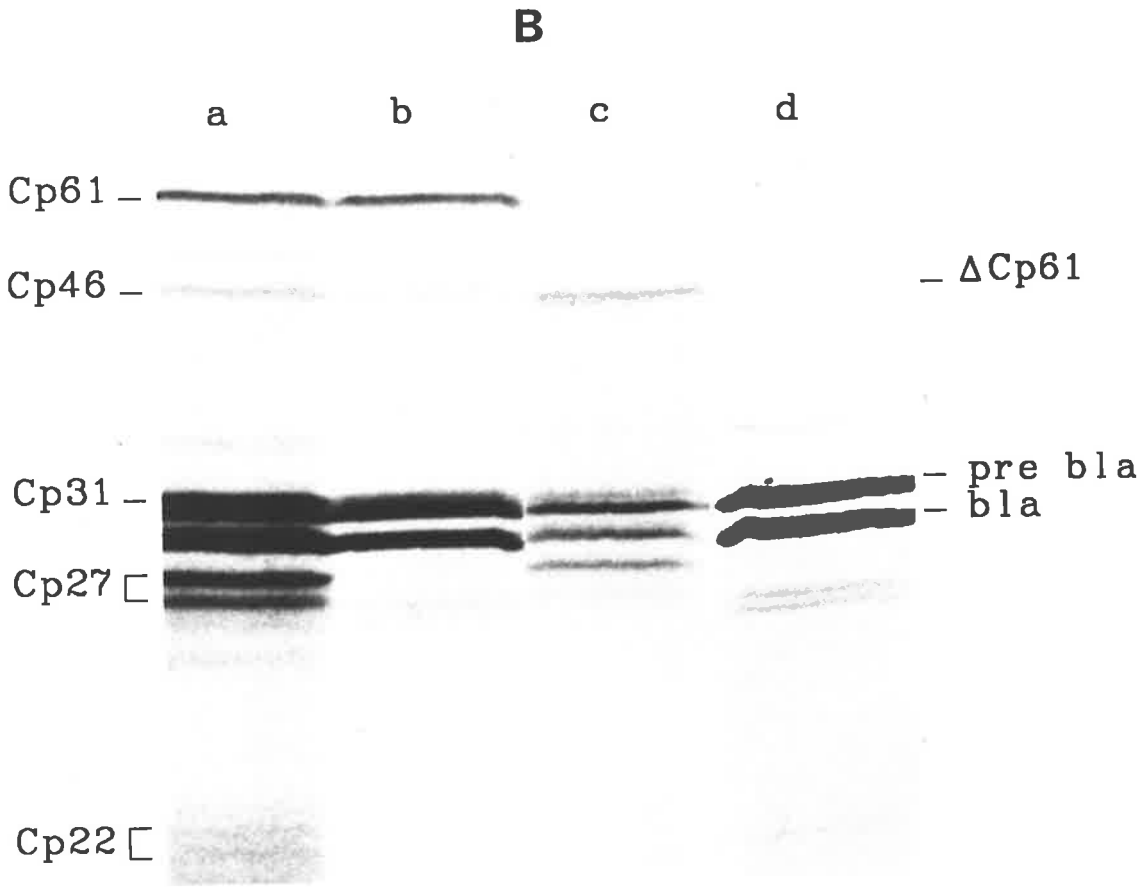
Lanes:

a - pPM484

b - The *Eco*RI mutant (pPM1353)

c - The *Bgl*II mutant (pPM1352)

d - pBR322


B
a
b
c
d
Cp61
Cp46
ΔCp61
Cp31
pre bla
bla
Cp27
Cp22

sequencing error (e.g. a nonsense codon being interpreted as sense codon) could be responsible for this. However, a critical examination of the sequence between the *Eco*RI site (nt 724), and the predicted termination site at nt 781, did not reveal any error. The observed disparity in mobility must therefore be due to an aberration of SDS-PAGE. The ratio of negatively charged amino acids (D+E) to positively charged ones (K+R) is 1.29 in the full length Cp27 polypeptide, whereas a ratio of 1.5 is seen in the truncated product. The relatively lower overall positive charge in the truncated product could be one reason for its aberrant migration.

4.2.1.2 Cp82

Although no plasmid-encoded protein band could be visualised in the vicinity of 82 kDa, two relatively intense bands of approximately 97 kDa and 94 kDa are consistently seen in minicell labelling of pPM484 (Figure 4.1, lane *c*), but not of pBR322. Previously reported by Manning *et al.* (1985), these bands are assumed to represent the *cp82* products. They were not detected in the *in vitro* translation of pPM484 (Figure 4.1, lane *b*).

End-filling of the unique *Nco*I site at nt 1663 (termed the *Nco*I mutation) resulted in a frameshift in the reading frame of *cp82* and caused the disappearance of the 97 kDa and 94 kDa bands, but did not affect any other (Figure 4.3A, lane *f*). A truncated product of 14.5 kDa, predicted from the sequence, was not found. The three AUG triplets at nt 1312, 1315 and 1318 are the only alternative initiation sites upstream to the *Nco*I site. The triplet at nt 1318 was accepted, taking into consideration that the other two AUG triplets would be too close to the putative RBS (AAGA at nt 1308). As in the case of *cp27*, a shift in the codon usage plot from poor to good coding potential occurred near this AUG (Figure 3.5). The protein bands of Cp61, Cp46, Cp31 and Cp9 did not show any change in their size or intensity, suggesting that a functional Cp82 was not required for their transcription and/or translation. Furthermore, it also proved that the bands assigned to their polypeptides were not derived from the former by post-translational

processing.

4.2.1.3 Cp61

A band of approximately 61 kDa was consistently seen in minicell labelling (Figure 4.1, lane *c*). Unlike those of Cp27 and Cp82, this band was not accompanied by another band. The predicted MW of the precursor protein is 61171 and, if it has a leader peptide of 20 aa, the mature peptide should have a MW of 58839. Owing to the absence of a MW marker closer to either of these sizes, it could not be said with certainty which of these bands was being observed. Presence of a single polypeptide signifies either an absence of signal peptide processing in minicells, or a rapid turnover of either the precursor or mature polypeptide.

The *Nco*I mutation had no effect on the *cp61* polypeptide (Figure 4.3A, lane *f*). However, a frameshift introduced by removal of the 3' overhang of the *Pst*I site at nt 2087 (termed the *Pst*I mutation) led to its disappearance (Figure 4.3A, lane *k*). Although a truncated product of 13.6 kDa was predicted from the sequence, this was not seen in minicell translation. As a further check to confirm the translational initiation site of *cp61*, an inframe deletion was effected by deleting a fragment between the two *Bgl*II sites (nt 2239 and 2587; termed the *Bgl*II deletion). This deletion eliminated Cp61 and produced a truncated product of 46 kDa (Figure 4.3A, lanes *c* and *j*). Since the only probable translational start sites between the *Nco*I (nt 1663) and *Pst*I (nt 2087) are the GUG and AUG triplets at nt 1837 and 1858, respectively, one of these must be the initiation codon of *cp61*. The AUG triplet at nt 1858 (Figure 3.4a) was taken as the more likely initiation site, although no suitable RBS was present before this codon.

It had been predicted from the sequence that the *Bgl*II deletion would truncate Cp61 to a 48 kDa precursor and, if it has a 20 aa leader sequence, to a 45.8 kDa mature peptide. As the observed band co-migrated with that of Cp46 (MW 45861) it appears to be the mature peptide and, by correlation, the full length product seen in the pPM484 lane must also be the mature protein of Cp61. However, no signal peptide could be

identified by examining its amino acid sequence. An alternative explanation is that the bands seen in pPM484 and the *Bgl*II mutant translations represent the precursor molecules which, due to abnormal migration on SDS-PAGE, appear shorter than their predicted sizes.

4.2.1.4 Cp46

A single band in the vicinity of 46 kDa (Figure 4.1, lane *c*) has been assigned to Cp46 (Manning *et al.*, 1985). In minicell translations this band appeared fainter as compared to those of Cp27 and CS3. The shorter band accompanying it in Figure 4.1 is probably an artefact, as in other experiments the 46 kDa polypeptide has been clearly observed as a single band (Figure 4.3A, lanes *a* and *i*).

Two alternative initiation sites for the product of *cp46* are the AUG triplets at nt 2131 and 2266, producing two polypeptides of 51 kDa and 46 kDa, respectively. The site-specific mutations generated in the analysis of *cp*61 proved useful in establishing the actual start site of its translation. Specifically, it was found that end-filling of the *Pst*I site at nt 2087 did not affect Cp46, which indicated that its initiation was downstream to this site. The *Bgl*II deletion, however, led to its disappearance (Figure 4.3A, lane *c* and *j*). A truncated product of approximately 38 kDa should have been present if translation initiated from the AUG at nt 2131. Absence of such a band suggested that its initiation site was located between these two *Bgl*II sites. The AUG triplet at nt 2266 was the only possible initiation codon in this DNA segment and, therefore, it was accepted as the start of *cp46*.

4.2.1.5 Cp31

A band migrating slightly above that of pre-β-lactamase (Pre-bla) has been observed in pPM484 translations. Failure to resolve these two bands from each other, in earlier experiments, had resulted in this polypeptide being overlooked (Manning *et al.*, 1985). It appeared to be produced at almost the same level as Cp27. In Figure 3.3, lane *a*, it migrates above that of Pre-bla, whereas lane *c* in Figure 4.1 displays a band below

that of Pre-bla. The absence of a signal sequence makes it difficult to assume that these represent the precursor and mature forms of Cp31. More likely, they represent the same form, which migrates differently under different electrophoresis experiments. Slight variations in SDS concentrations can affect the relative migration of proteins (Manning and Reeves, 1977).

The two alternative initiation sites for *cp31* are the AUG triplets at nt 291 and 2668, respectively, which would give either a 30.4 kDa or 30.7 kDa polypeptide. The first potential start at nt 291 was ruled out when it was found that the end-filling of *Eco*RI site at nt 724 did not result in the disappearance of the 31 kDa band (Figure 4.3B, lane *b*). The *Bgl*II deletion (nt 2239 and 2587) also did not result in its disappearance (Figure 4.3B, lane *c*), suggesting that it originated downstream to nt 2587. The only possible start site for *cp31* was therefore the AUG at nt 2668.

4.2.1.6 Cp22

Two bands in the vicinity of 22 kDa are seen in minicell (Figure 4.2, lane *c*) as well as *in vitro* translations of pPM484 (Figure 4.1, lane *b*, Figure 4.4, lane *a*). The shorter band accumulated when pilus biosynthesis was inhibited by the *Bgl*II deletion (Figure 4.3A, lane *j*).

Removal of UAG_{3523} (*see* section 4.2.2) by oligomutagenesis, and subsequent frameshift at the resulting *Sac*I site, did not affect the *cp22* polypeptide (Figure 4.5), which indicated that its coding region was downstream to this codon. One of the two *Apa*I sites, at nt 3706 and 4226, is in the coding region of *cp22*. Deleting the fragment between these *Apa*I sites (termed the *Apa*I deletion) showed the disappearance of the two bands assigned to Cp22 (Figure 4.4, lanes *b* and *c*). Although these two bands possess a relative size difference of about 2 kDa, suggesting that they represent the pre- and mature forms of the same protein, their appearance in the *in vitro* translation rules out this possibility. More likely, these bands originate from alternative initiations of translation. Although initiation from the AUG at nt 3530 could give a 22 kDa band, no

Figure 4.4 Effect of *Apa*I deletion on Cp22 and CS3.

In vitro and *in vivo* (minicell) translation of pPM484 and its derivaive carrying a deletion of the fragment between nt 3706 and 4226. The proteins were labelled with ^{35}S-Met and separated in SDS on 15% PAGE. Band positions are indicated. Also indicated, in brackets, are the expected positions of protein products not visible in the gels.

Lanes:

a - *In vitro* translation of pPM484

b - *In vitro* translation of the *Apa*I mutant (pPM4017)

c - *In vivo* translation of the *Apa*I mutant (pPM4017)

d - *In vivo* translation of pPM484

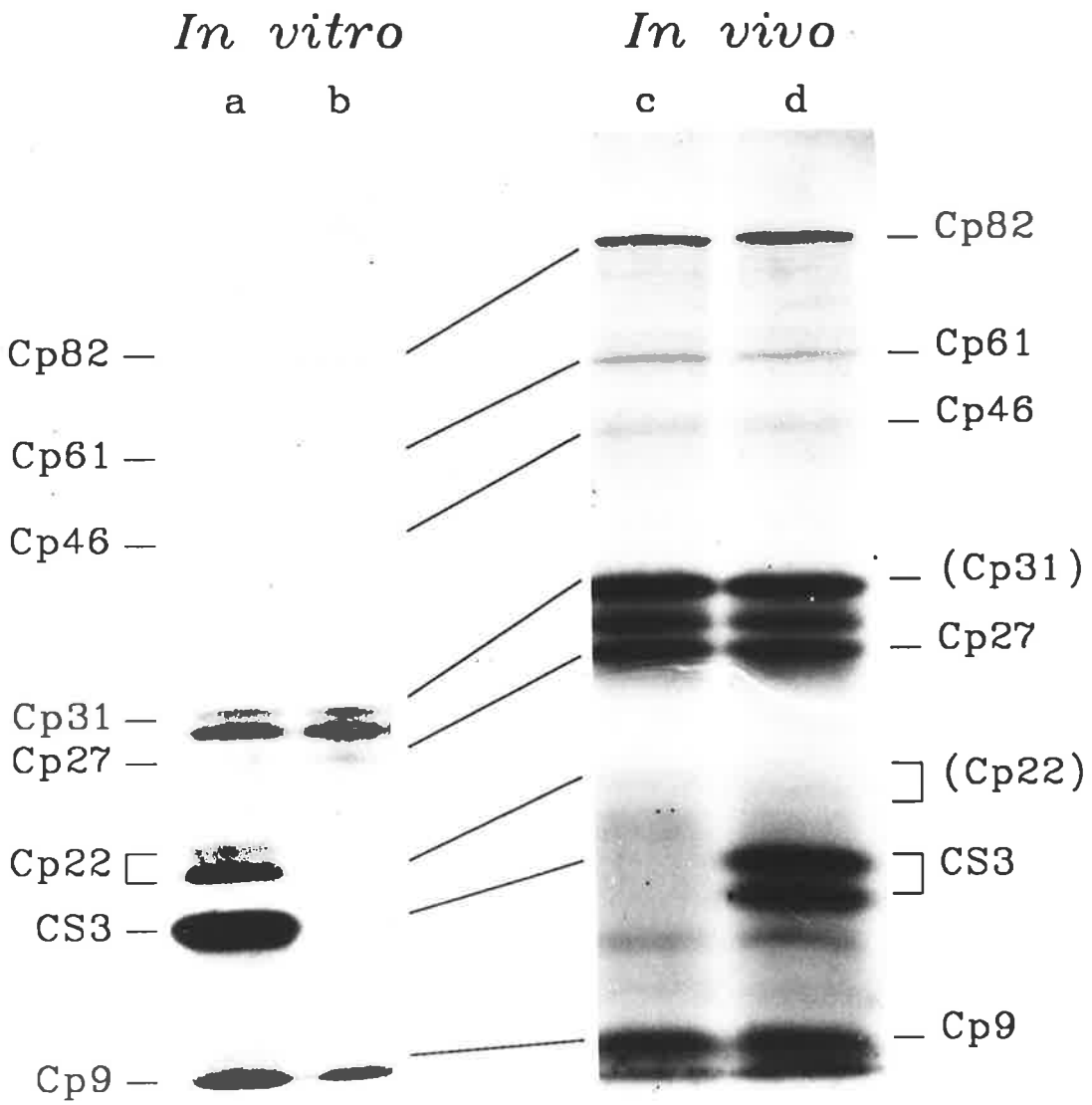

In vitro
In vivo
a
b
c
d
Cp82
Cp61
Cp46
Cp31
Cp27
Cp22
CS3
Cp9
Cp82
Cp61
Cp46
(Cp31)
Cp27
(Cp22)
CS3
Cp9

Figure 4.5 Minicell labelling of the UAG_{3523} mutant

In vivo (minicell) translation products of pPM484 and its mutant derivatives were run in SDS on 15% PAGE.

Lanes:

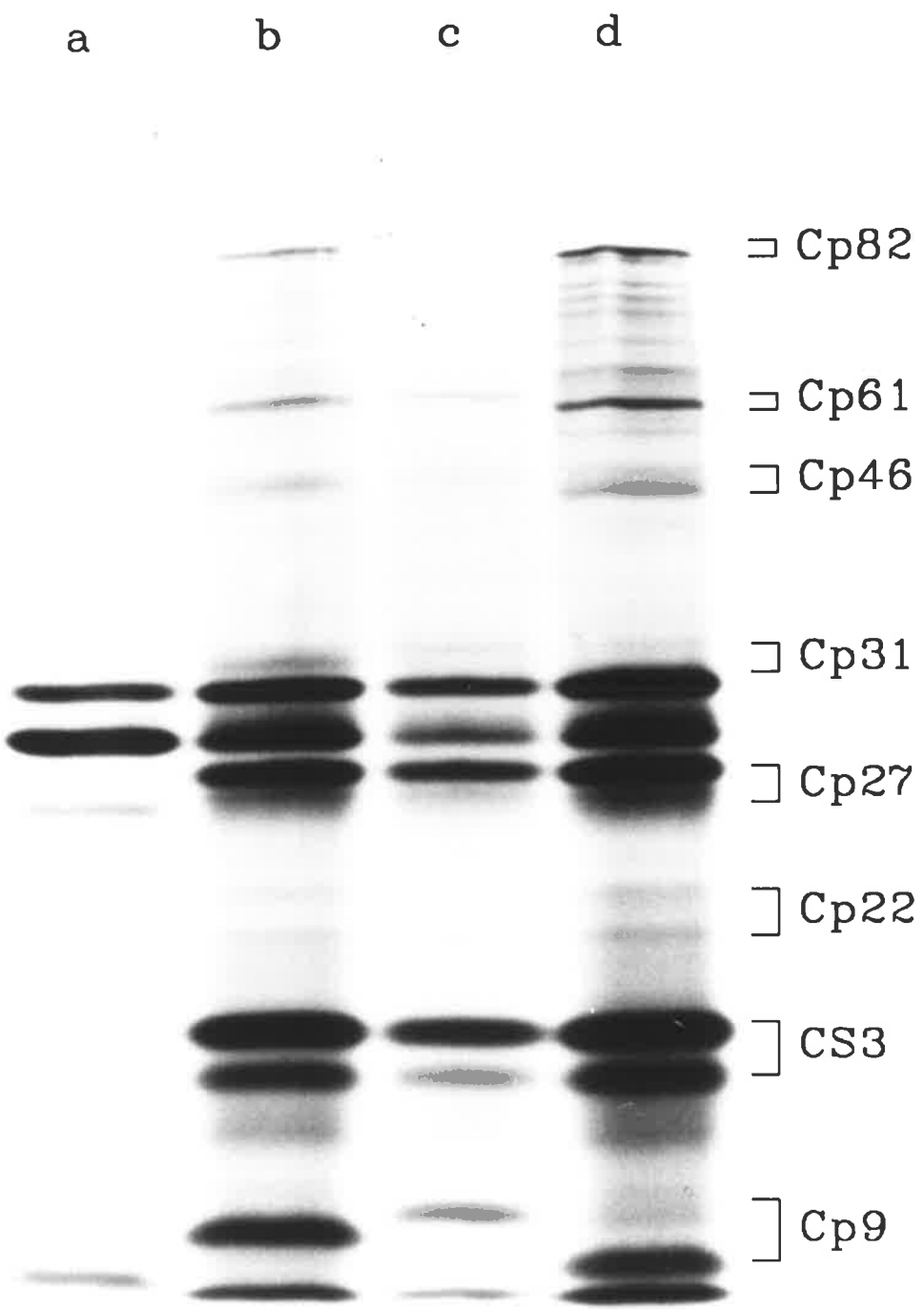
a - pBR322

b - The *Sac*I mutant

c - The UAG_{3523} mutant

d - pPM484

The normal and extended products for Cp82, Cp61, Cp46, Cp31 and Cp9 are indicated. Alsó indicated are the precursor and mature forms of Cp27, Cp22 and CS3.


a
b
c
d
Cp82
Cp61
Cp46
Cp31
Cp27
Cp22
CS3
Cp9

potential initiation codon could be assigned for the shorter band.

The translation of *cp22* is unusual, requiring an apparent suppression of the amber codon, UAG_{3578}. The observation that CS3 pilus formation does not occur in a suppressor-free strain of *E. coli* C (*see* section 6.2.1) lends support to this assumption.

4.2.1.7 Cp9

Although often overlooked, a 9 kDa band is seen in minicell translations (Figure 3.3, lane *b,* Figure 4.5, lane *d*). The intensity of this band suggests that it is made in larger amounts than those of Cp46 and Cp61. *In vitro* translation also reveals this band (Figure 4.1, lane *b*).

The absence of a suitable restriction site within its coding region made site-directed mutagenesis of *cp9* difficult. However, mutagenesis of UAG_{3523} (*see* section 4.2.2), and subsequent frame-shift, resulted in extended polypeptides of Cp9 as predicted from the sequence (Figure 4.5, lanes *b* and *c*). Owing to the relatively small size of the coding region, the initiation site was deduced by searching for an AUG triplet at an appropriate distance from UAG_{3523}. Only one potential start site, the AUG triplet at nt 3274, could produce a polypeptide of 9 kDa (predicted MW of 8668).

4.2.1.8 CS3

Two bands of approximately 15 and 17 kDa are seen in minicell translations (Figure 4.1, lane *c*). The larger of these represent the precursor form, as it alone is present in *in vitro* translations (Figure 4.1, lane *b*, Figure 4.4, lane *a*). The smaller band was detected by Western blotting using antisera raised against the purified pili from wild type ETEC strain, PB176, and corresponds in length to the minor form of the latter (data not shown).

The N-terminal amino acid sequence of the first 30 residues, determined from CsCl-purified CS3 pili from original wild type ETEC strain PB176, corresponded exactly with that for amino acids 23-52 of the *cpCS3* product. The preceding 22 amino acids constitute a signal sequence (Figure 3.4a), according to the rules of von Heijne (1984,

1985, 1986). Since the mature polypeptide would be devoid of any Met residue, it should not have been labelled with ^{35}S-Met in minicell translations. However, labelled bands corresponding to both the precursor and apparent mature forms could be seen in minicell translations. Analysis of the N-terminal amino acid sequence data revealed a second product in lower molar ratio corresponding to an alternative cleavage between residues 14 and 15 (Ser and Ala). In this case the mature polypeptide is longer, and would retain one Met residue. This cleavage would fit in with the -1, -3 rule (von Heijne 1984, 1985, 1986), although the h- region is at the lower limits of size. It does, however, provide a helix-breaker (Gly) at -5 from the cleavage site. The banding pattern of purified pili from PB176 and *E. coli* K-12 differed, in that the former showed two bands. Only the minor, larger band was present in *E. coli* K-12 (Figure 3.1).

The AUG triplet at nt 4154, preceded by a suitable RBS, is assigned to be the initiation codon of CS3. Presence of a potential promoter between nt 4107 and 4135 also suggested that it was the probable candidate. A deletion of the DNA between the two *Apa*I sites at nt 3706 and 4226 showed the disappearance of the CS3 polypeptides (Figure 4.4, lanes *b* and *c*).

4.2.2 Termination codon of the overlapping ORFs

Five ORFs (*cp82*, *cp61*, *cp46*, *cp31* and *cp9*) terminated at a common UAG triplet at nt 3523. Oligo-nucleotide directed mutagenesis of this stop codon to a sense codon should result in lengthening of the respective polypeptides by 24 amino acids (ca. 2.6 kDa), as translation would extend up to the UAA at nt 3595. Substitution of G in place of T at nt 3523 resulted in the removal of the stop codon and creation of a unique *Sac*I site at this site (termed the UAG_{3523} mutation). The mutation was verified by *Sac*I restriction analysis of plasmid DNA (Figure 4.6, lanes *c* and *h*). As a further verification of the introduced change, as well as fidelity of the DNA sequence between the two UAG triplets (nt 3523 and 3578), a frameshift was introduced by removal of the 3' overhang generated by this enzyme and the change (termed the *Sac*I mutation) was verified by

Figure 4.6 Restriction analysis to verify the UAG_{3523} mutation.

Plasmid DNAs of pPM484 and its derivatives, digested with different restriction enzymes, were run on 1% agarose, stained with ethidium bromide and photographed under short wave UV.

Lanes and expected sizes of bands, in brackets, are:

a - pBR322 x *Hin*dIII (4.36 kb)

b - pPM484 x *Nco*I (9.1 kb; control for linear pPM484)

c - UAG_{3523} mutant x *Sac* I (9.1 kb)

d - pPM484 x *Hin*dIII (4.36 and 4.75kb)

e - pPM484 x *Hin*dIII + *Sac*I (4.36 and 4.75 kb)

f - pPM484 x *Sac*I (uncut supercoiled DNA)

g - UAG_{3523} mutant x *Hin*dIII (4.36 and 4.75kb)

h - UAG_{3523} mutant x *Hin*dIII + *Sac*I (4.36, 3.52 and 1.23 kb)

i - SPP1 x *Eco*RI (Size standards; Bresatec, Adelaide)

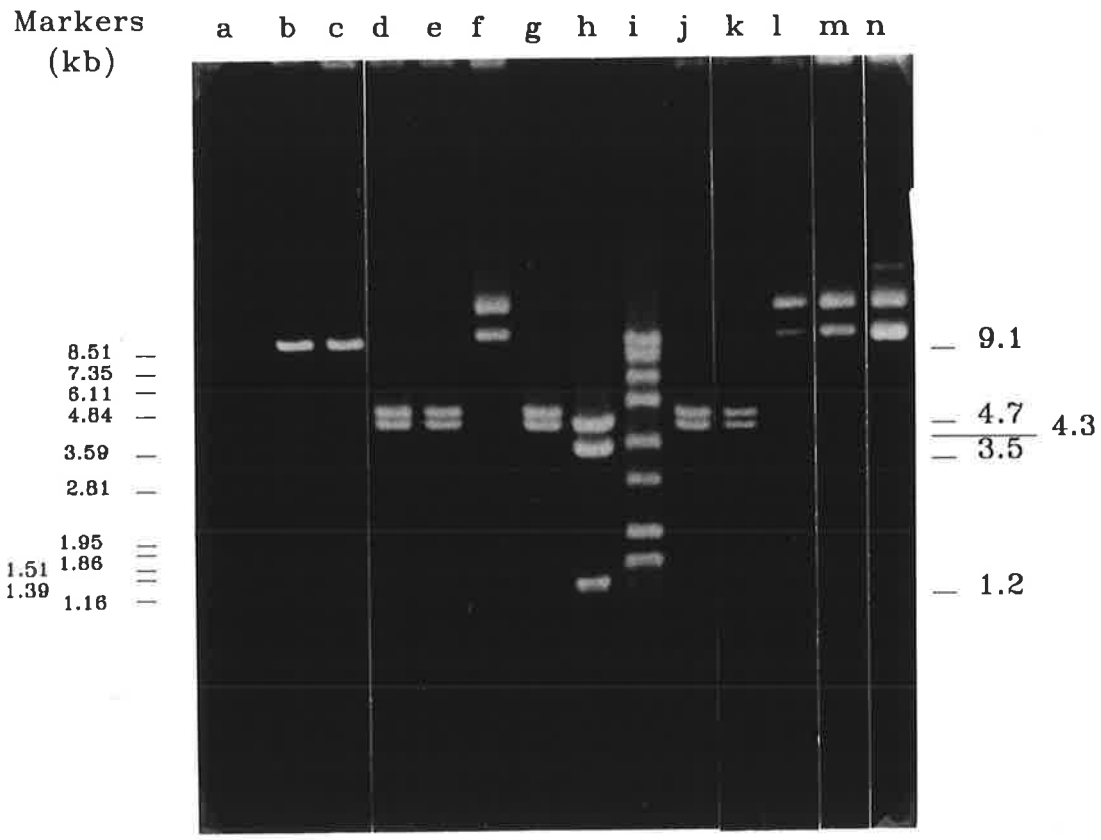
j - *Sac*I mutant x *Hin*dIII (4.36 and 4.75kb)

k - *Sac*I mutant x *Hin*dIII + *Sac*I (4.36 and 4.75 kb)

l - *Sac*I mutant x *Sac*I (uncut supercoiled DNA)

m - uncut UAG_{3523} mutant

n - uncut pPM484.


Markers
(kb)
a b c d e f g h i j k l m n
8.51
7.35
6.11
4.84
3.59
2.81
1.95
1.86
1.51
1.39
1.16
9.1
4.7
4.3
3.5
1.2

restriction analysis (Figure 4.6, lanes *c, h, k* and *l*). The ORFs now terminated at the UAG triplet at nt 3578, adding only 17 aa (1.9 kDa) to the wild type polypeptides. Minicell translations of the two mutant plasmids revealed that the products of *cp82, cp61, cp46, cp31,* and *cp9* increased in sizes as predicted (Figure 4.5, lanes *b* and *c*). No change in size was observed for the protein bands assigned to Cp27, Cp22 and CS3.

4.2.3 Involvement of the Cp proteins in pilus biosynthesis.

The overlapping arrangement of the genes posed a difficulty in ascertaining the roles of individual genes in pilus biosynthesis. Only the mutagenesis of *cp27* and *cp82* could be successfully achieved without a concomitant effect on other genes. Direct involvement of only these two genes could therefore be verified, and the involvement of others is assumed in the following discussion.

4.2.3.1 Cp27

The effect of the *Eco*RI mutation on pilus biosynthesis was verified by western blotting of crude pili (Figure 4.7, lane *c*) and electron microscopic examination by immunogold labelling (Figure 4.8b). No pili were detected in either case. Minicell analysis showed the presence of the CS3 subunit (Figure 4.3A, lane *b*), which indicated that functional Cp27 was not needed for the transcription and/or translation of CS3 subunit. However, it is clearly required for the assembly of cell-surface pili.

4.2.3.2 Cp82

The *Nco*I mutation resulted in an absence of cell-surface pili, as indicated by Western blotting of crude pili preparation (Figure 4.7, lane *d*) and immunogold EM (Figure 4.8c). It was shown by ^{35}S-Met labelling that the CS3 subunit was being produced in minicells (Figure 4.3A, lane *f*). This indicated that functional Cp82 was not required for the expression and post-translational modifications, if any, of the CS3 polypeptide. However, the absence of detectable pili indicated that it was needed for the formation of cell-surface pili. There was no phenotypic difference between cells carrying

Figure 4.7 Western blot of pili from cells carrying the mutant plasmids

Crude pili, prepared from cells carrying the different mutant derivatives of pPM484, were separated in SDS on 15% PAGE, transferred to nitrocellulose and probed with CS3 antiserum. The CS3 subunit is indicated. Other visible bands do not correspond to any of the pPM484-specific proteins and probably are the results of non-specific antibody reaction.

Lanes:

a - pPM484

b - pBR322

c - The *Eco*RI mutant

d - The *Nco*I mutant

e - The *Bgl*II deletion mutant

f - The *Pst*I mutant

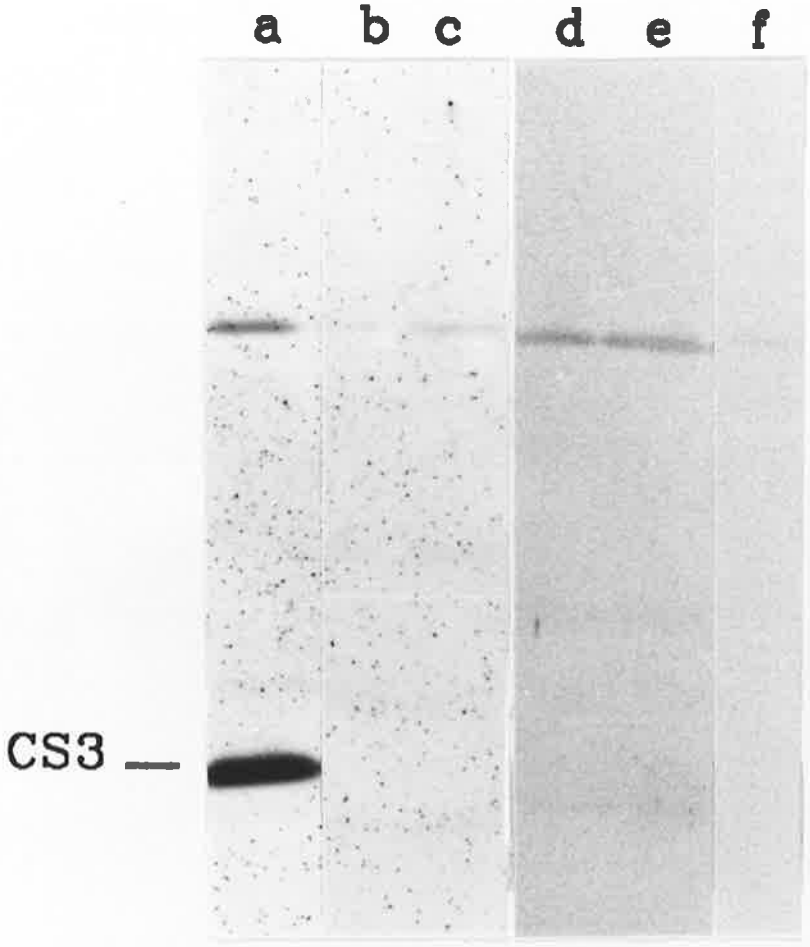

a
b
c
d
e
f
CS3

Figure 4.8 Immunogold EM of cells carrying mutant derivatives of pPM484

Minicells carrying pPM484 or its mutant derivatives were labelled with rabbit anti-CS3 pilin antibody, followed by protein-A colloidal gold complexes. Photographs are taken at different magnifications. Thus, horizontal bar equals 0.14μM in *a*, but 0.37μM in all other panels.

a - pPM484

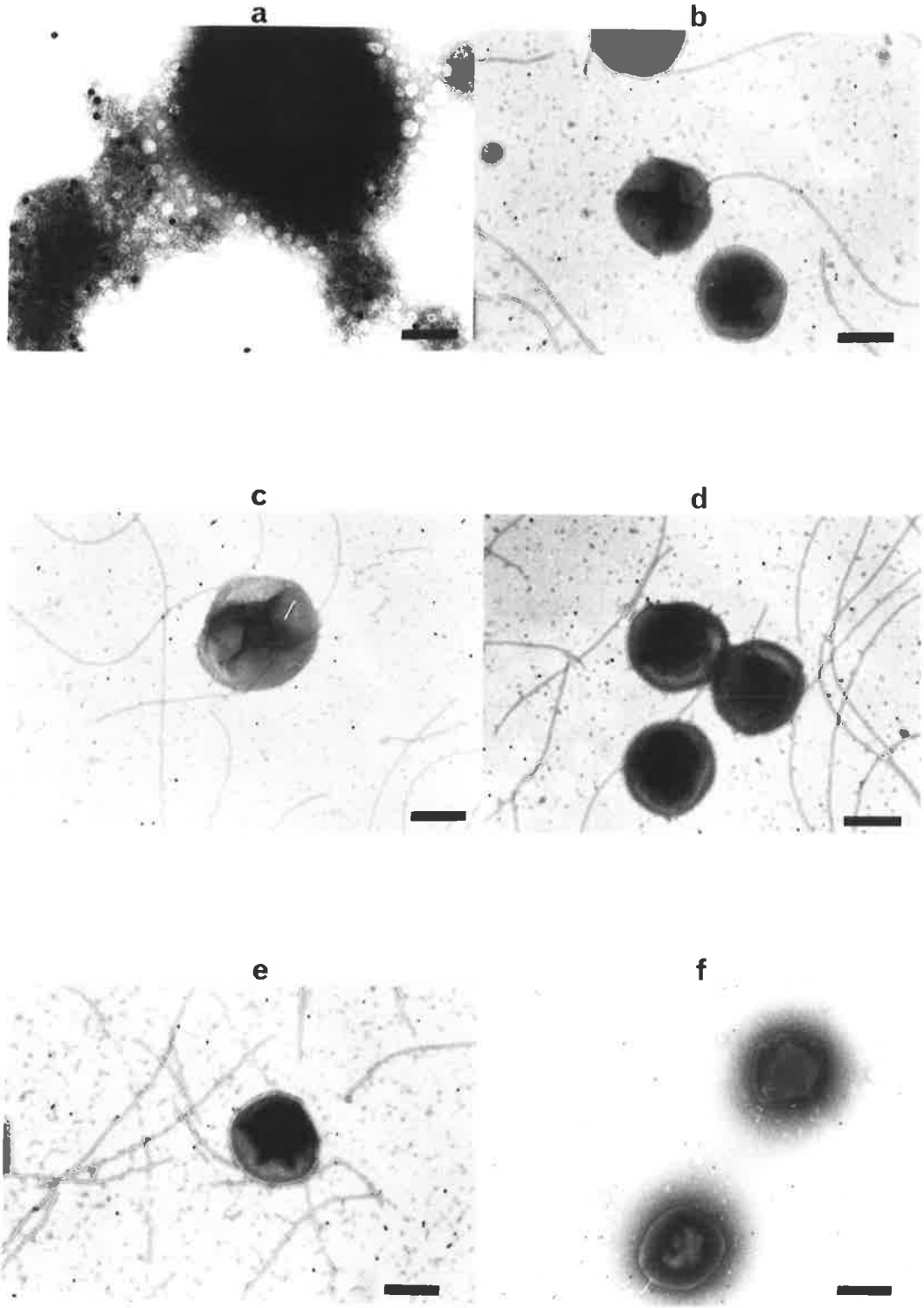
b - *Eco*RI mutant

c - *Nco*I mutant

d - *Bgl*II deletion mutant

e - *Pst*I mutant

f - pBR322


a
b
c
d
e
f

the *cp82*-mutated plasmid and those carrying the mutation in *cp27*.

4.2.3.3 Cp61 and Cp46

Western blotting showed that the *Bgl*II deletion, which truncated the product of Cp61 and eliminated that of Cp46, resulted in an absence of CS3 pili on the cell surface (Figure 4.7, lane *e*, Figure 4.8d). Similarly, the end-filling of the *Pst*I site nt 2087 resulted in absence of cell surface pili, as evidenced by Western blot and immunogold EM (Figure 4.7, lane *f*, Figure 4.8e). Although the absence of assembled pili indicated that the products of these genes were needed for its formation, both these mutations also affected the product of Cp82. Since a mutation in *cp82* was associated with an absence of pili, and *cp82* overlapped with *cp61* and *cp46*, the effect could have been due to their effect on the Cp82 protein. A direct role of Cp61 and Cp46 could not be ascertained from this experiment, however, the CS3 band in minicell labelling indicated an absence of their involvement in the transcription and/or translation of *cpCS3* (Figure 4.3A, lanes *c, j* and *k*).

4.2.3.4 Cp31 and Cp9

The *cp31* and *cp9* ORFs have not been analysed by site-specific mutagenesis. However, read-through translation of these, effected by mutagenising the UAG_{3523} resulted in loss of pilus synthesis as indicated by immunogold EM (*see* Figure 6.6c). In this case, however, a very faint band corresponding to CS3 subunit was seen in Su^+, but not in Su^-, background by Western blotting (*see* Figure 6.3, lanes *g* and *h*). Again, as in the case of Cp61 and Cp46, this could reflect an effect of the change in Cp82. Interestingly, end-filling of the newly created *Sac*I site at UAG_{3523} returned the pilus synthesis capacity to the cells (*see* Figure 6.3, lanes *e* and *f* ; Figure 6.6d). It appears that the addition of a certain string of amino acids to the C-terminal end of one or more proteins affects pilus synthesis, whereas addition of another string does not.

4.2.3.5 Cp22

Although one of the two *Apa*I sites occurred within the coding region of Cp22, attempts to end-fill this site without affecting the other were unsuccessful. Therefore, direct evidence for the involvement of this protein in pilus synthesis could not be obtained. However, it has been observed that pili are not formed in an Su^- minicell strain (section 6.2.1). This effect is believed to be due to the need to suppress the inframe UAG_{3578} codon within the reading frame of *cp22*. If so, it can be inferred that *cp22* is involved in pilus biosynthesis. Further evidence for its possible involvement as a structural component of the pilus itself is presented later in this chapter (*see* section 4.2.7).

4.2.3.6 CS3

Direct correlation of the N-terminal amino acid sequence of purified pili and the deduced amino acid sequence of *cpCS3* proved that it was the structural gene for CS3 subunit. No further mutational analysis was done on this ORF.

4.2.4 Structural properties of the proteins

The primary amino acid sequences of the various proteins were analysed by computer programs to predict their secondary structure properties. In the following sections the secondary structure predictions, by the method of Chou and Fasman (1978), and hydrophobicity by Kyte and Doolittle (1982) are described for each protein. The GCG programs, "peptidestructure" and "plotstructure" (Devereux *et al.*, 1984) are used to make these predictions, and the outputs for individual proteins are given in Figures 4.9 to 4.16.

4.2.4.1 Cp27

Table 4.1 lists the amino acid composition of Cp27. Figure 4.9 shows its protein secondary structure, superimposed with hydrophilicity and hydrophobicity. The N-terminal region is highly hydrophobic, characteristic of a leader peptide (Sjostrom *et al.*, 1987).

Table 4.1
Summary of protein compositions

Ptn	Cp27		Cp82		Cp61		Cp46		Cp31		Cp9		Cp22		CS3	
aa	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Cys	3	1.24	4	0.54	2	0.36	2	0.48	2	0.70	2	2.41	0	0.00	0	0.00
Ser	15	6.22	78	10.61	66	11.89	50	11.93	28	9.82	5	6.02	22	11.06	20	11.90
Thr	14	5.81	40	5.44	34	6.13	25	5.97	18	6.32	5	6.02	24	12.06	27	16.07
Pro	15	6.22	21	2.86	15	2.70	8	1.91	7	2.46	1	1.20	6	3.02	5	2.98
Ala	11	4.56	30	4.08	22	3.96	17	4.06	12	4.21	4	4.82	10	5.03	16	9.52
Gly	17	7.05	74	10.07	60	10.81	49	11.69	38	13.33	13	15.66	15	7.54	9	5.36
Asn	18	7.47	61	8.30	43	7.75	40	9.55	24	8.42	6	7.23	11	5.53	18	10.71
Asp	11	4.56	43	5.85	29	5.23	23	5.49	18	6.32	8	9.64	12	6.03	6	3.57
Glu	13	5.39	32	4.35	22	3.96	12	2.86	7	2.46	1	1.20	6	3.02	4	2.38
Gln	6	2.49	23	3.13	17	3.06	11	2.63	7	2.46	2	2.41	4	2.01	2	1.19
His	2	0.83	5	0.68	4	0.72	4	0.95	3	1.05	0	0.00	3	1.51	2	1.19
Arg	9	3.73	27	3.67	17	3.06	14	3.34	8	2.81	1	1.20	5	2.51	1	0.60
Lys	22	9.13	45	6.12	32	5.77	25	5.97	17	5.96	6	7.23	13	6.53	8	4.76
Met	4	1.66	11	1.50	9	1.62	7	1.67	6	2.11	3	3.61	4	2.01	2	1.19
Ile	24	9.96	54	7.35	44	7.93	34	8.11	27	9.47	8	9.64	13	6.53	11	6.55
Leu	25	10.37	53	7.21	40	7.21	32	7.64	17	5.96	5	6.02	16	8.04	21	12.50
Val	13	5.39	36	4.90	30	5.41	20	4.77	19	6.67	8	9.64	17	8.54	9	5.36
Phe	9	3.73	33	4.49	20	3.60	10	2.39	5	1.75	1	1.20	11	5.53	3	1.79
Tyr	7	2.90	54	7.35	42	7.57	29	6.92	18	6.32	3	3.61	5	2.51	3	1.79
Trp	3	1.24	11	1.50	7	1.26	7	1.67	4	1.40	1	1.20	2	1.01	1	0.60
Total	241	100	735	100	555	100	419	100	285	100	83	100	199	100	168	100
MW	26893		82150		61171		45861		30687		8668		21704		17475	

Figure 4.9 Secondary structure and hydrophobicity profiles of Cp27

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (○) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

labelling reversed

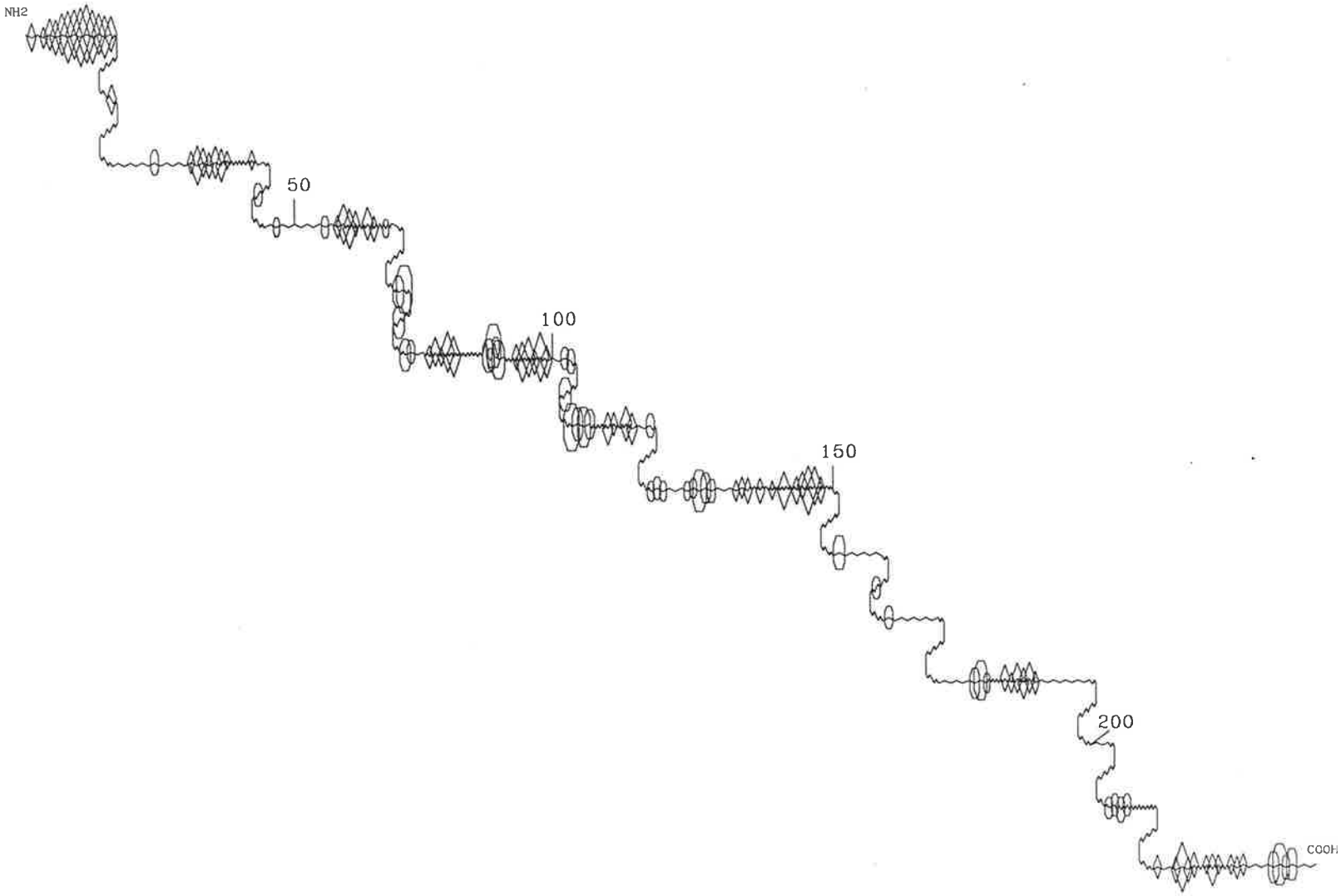

NH2
50
100
150
200
COOH

Figure 4.10 Secondary structure and hydrophobicity profiles of Cp82

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (⬡) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

Positions of the start sites of other proteins encoded within the coding region of Cp82 are marked.

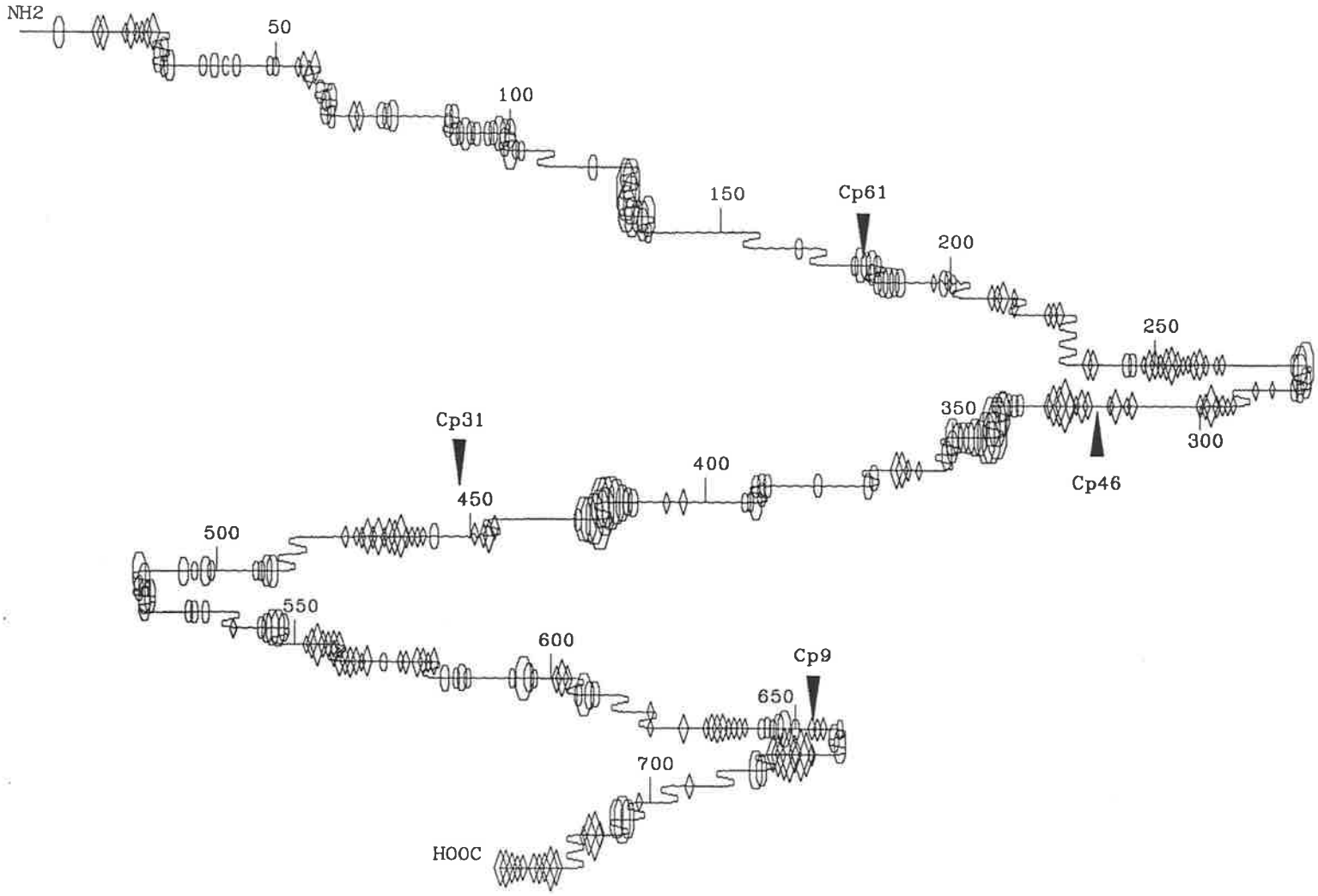

NH2
50
100
150
Cp61
200
250
300
350
Cp46
400
Cp31
450
500
550
600
Cp9
650
700
HOOC

Figure 4.11 Secondary structure and hydrophobicity profiles of Cp61

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (○) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

Positions of the start sites of other proteins encoded within the coding region of Cp61 are marked.

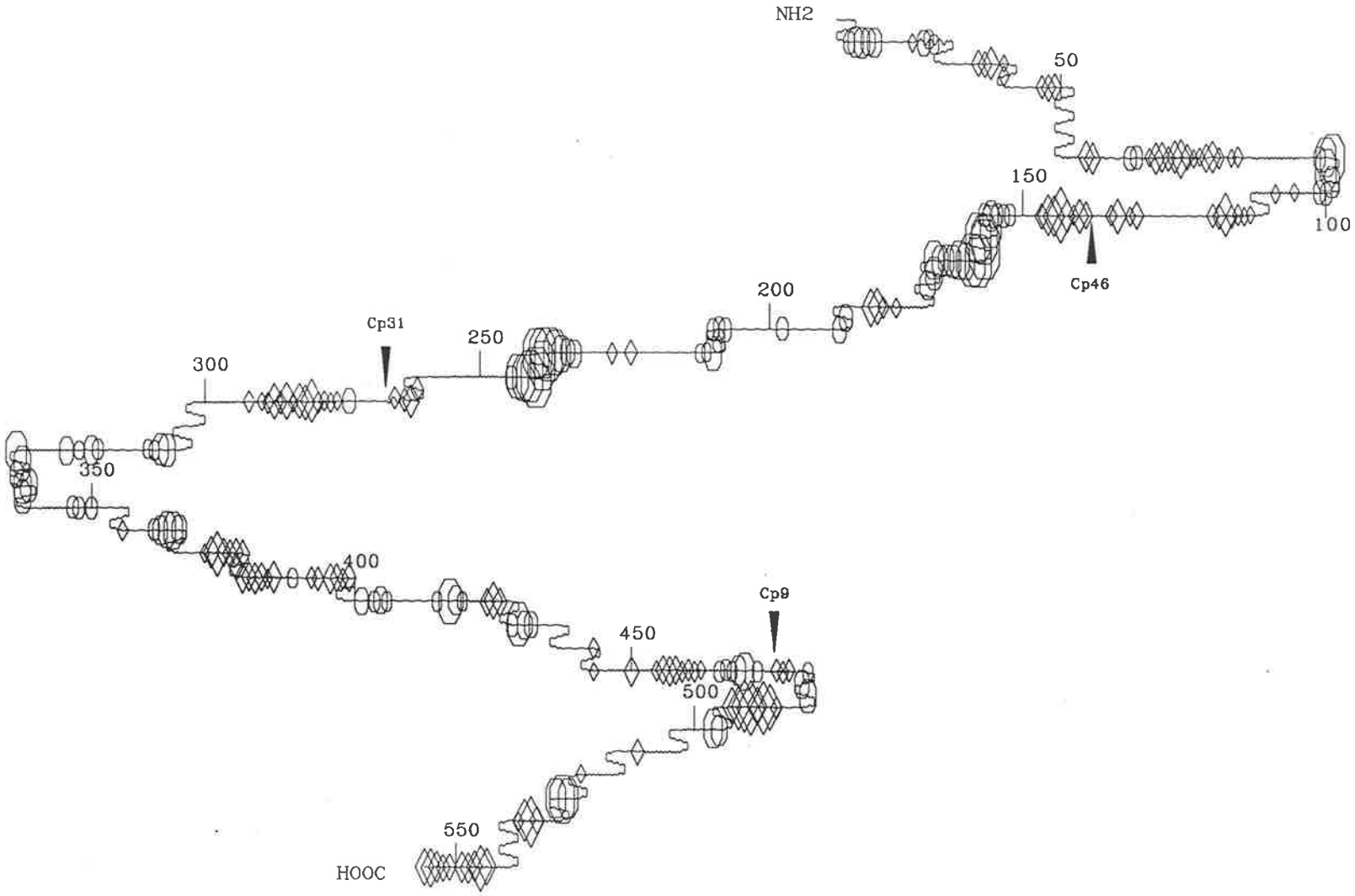

NH2
50
100
150
200
Cp46
Cp31
250
300
350
400
Cp9
450
500
550
HOOC

Figure 4.12 Secondary structure and hydrophobicity profiles of Cp46

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (○) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

Positions of the start sites of Cp31 and Cp9, encoded within the coding region of Cp46, are marked.

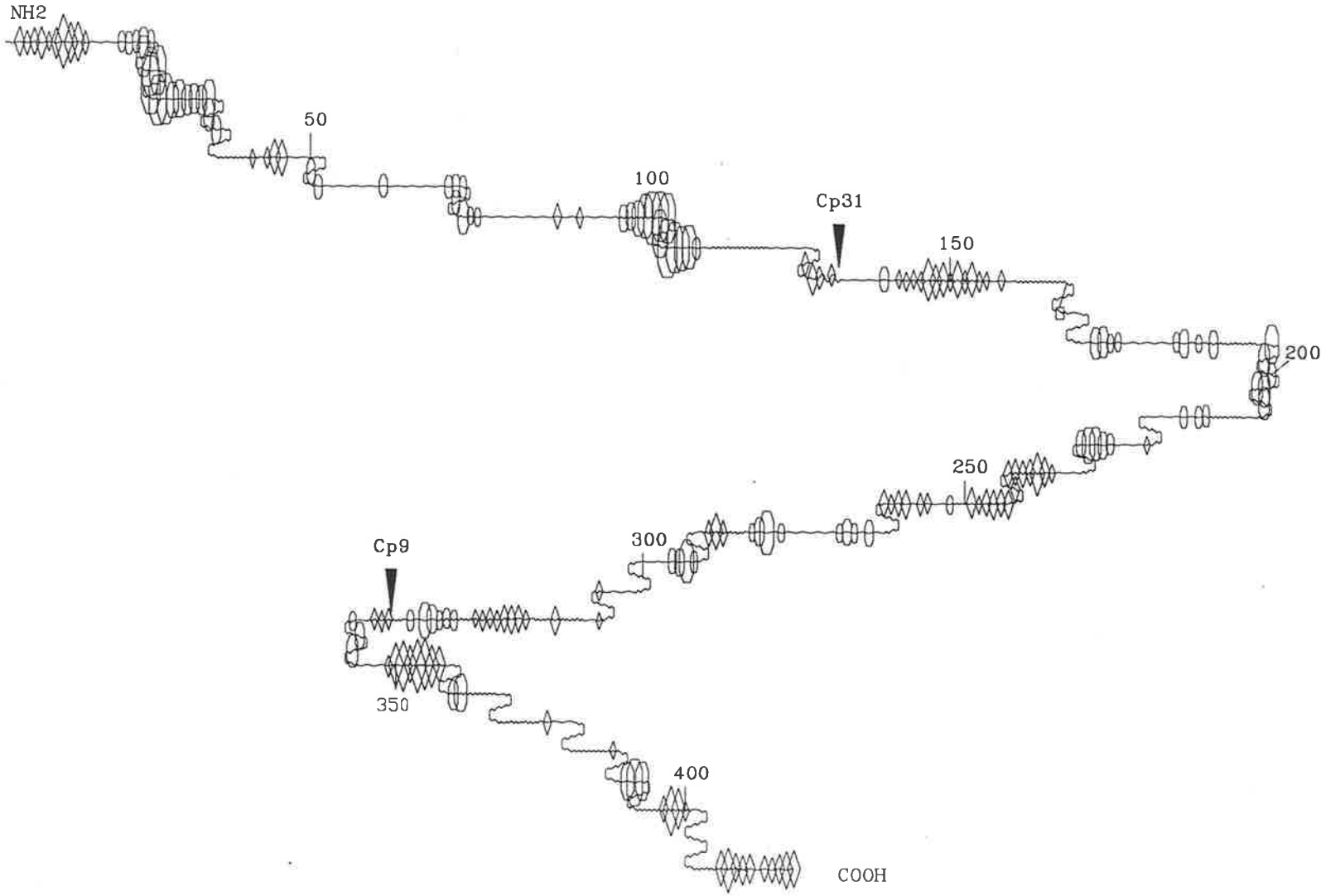

NH2
50
100
Cp31
150
200
250
300
Cp9
350
400
COOH

Figure 4.13 Secondary structure and hydrophobicity profiles of Cp31

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (▢) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

Position of the Cp9 start site within the coding region of Cp31 is marked.

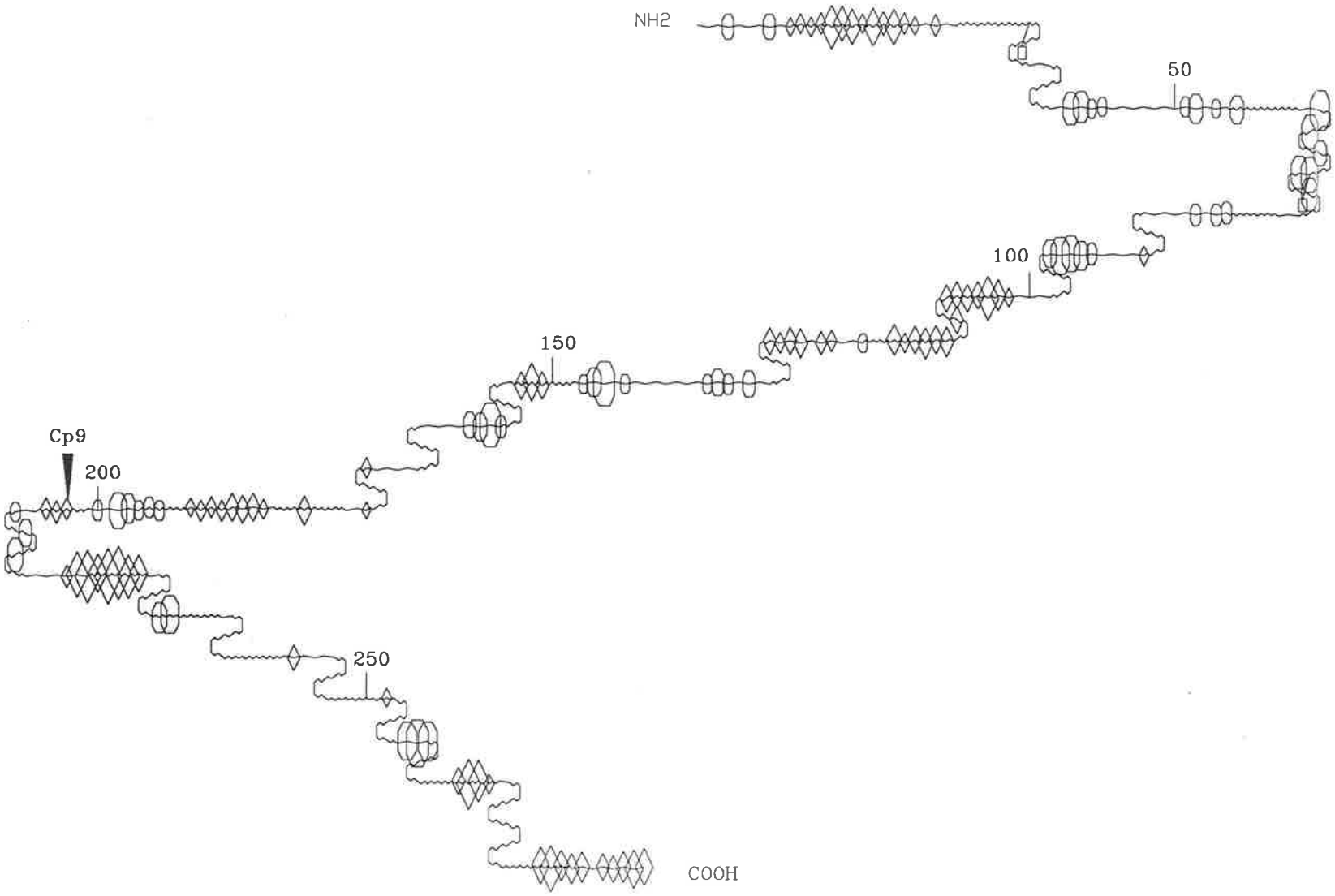

NH2
50
100
150
Cp9
200
250
COOH

Figure 4.14 Secondary structure and hydrophobicity profiles of Cp9

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (⬡) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

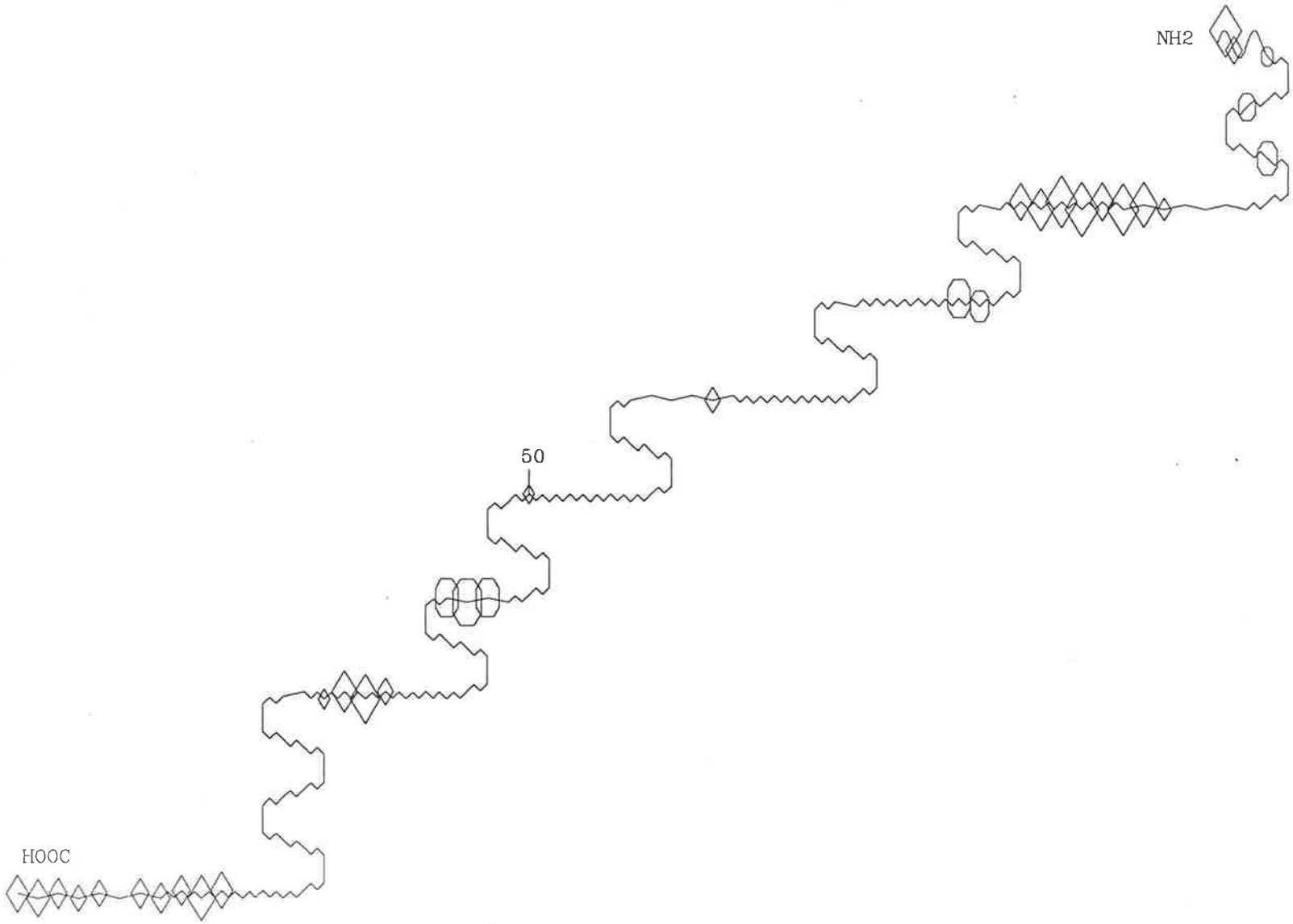

NH2
50
HOOC

Figure 4.15 Secondary structure and hydrophobicity profiles of Cp22

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (○) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).

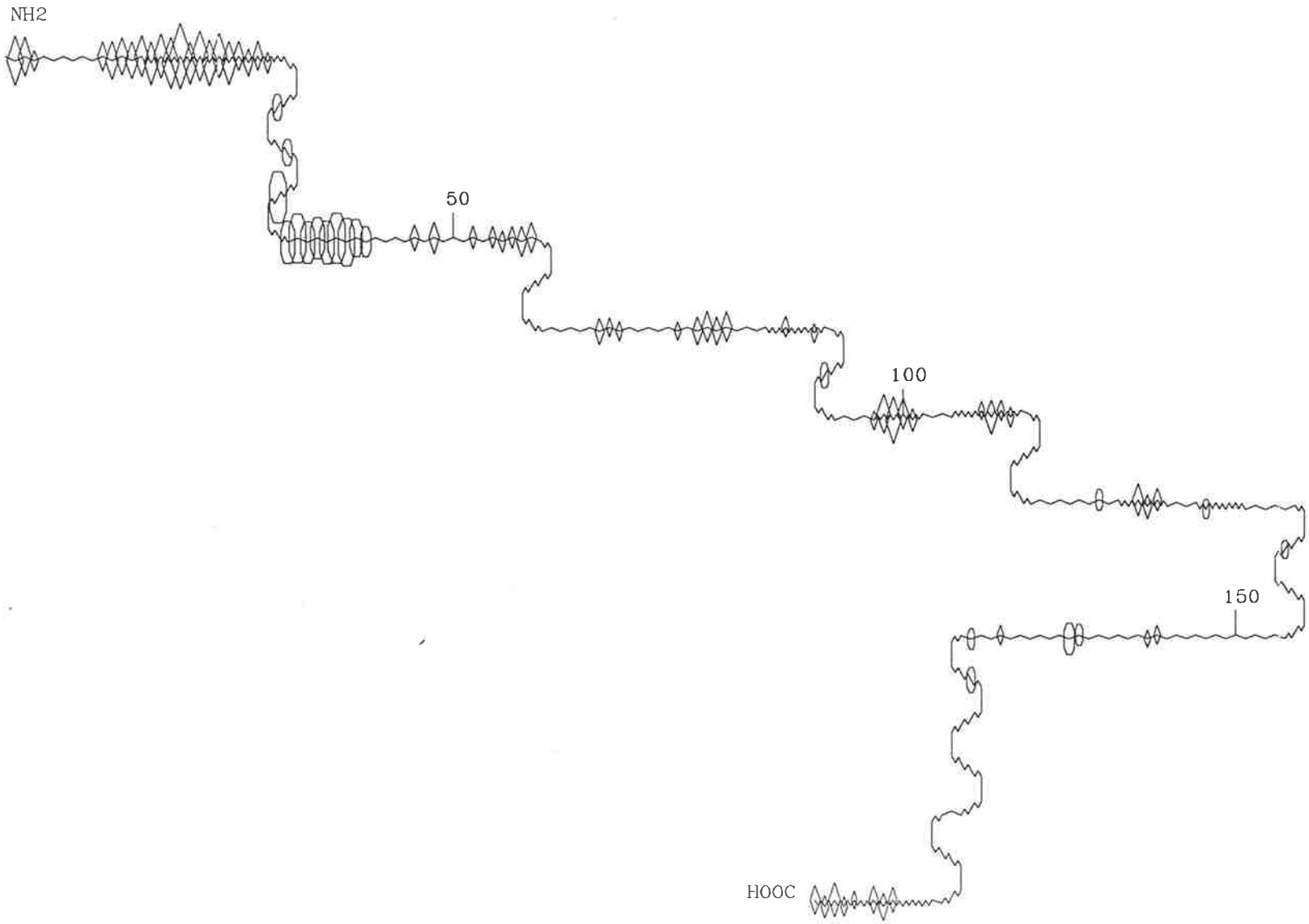
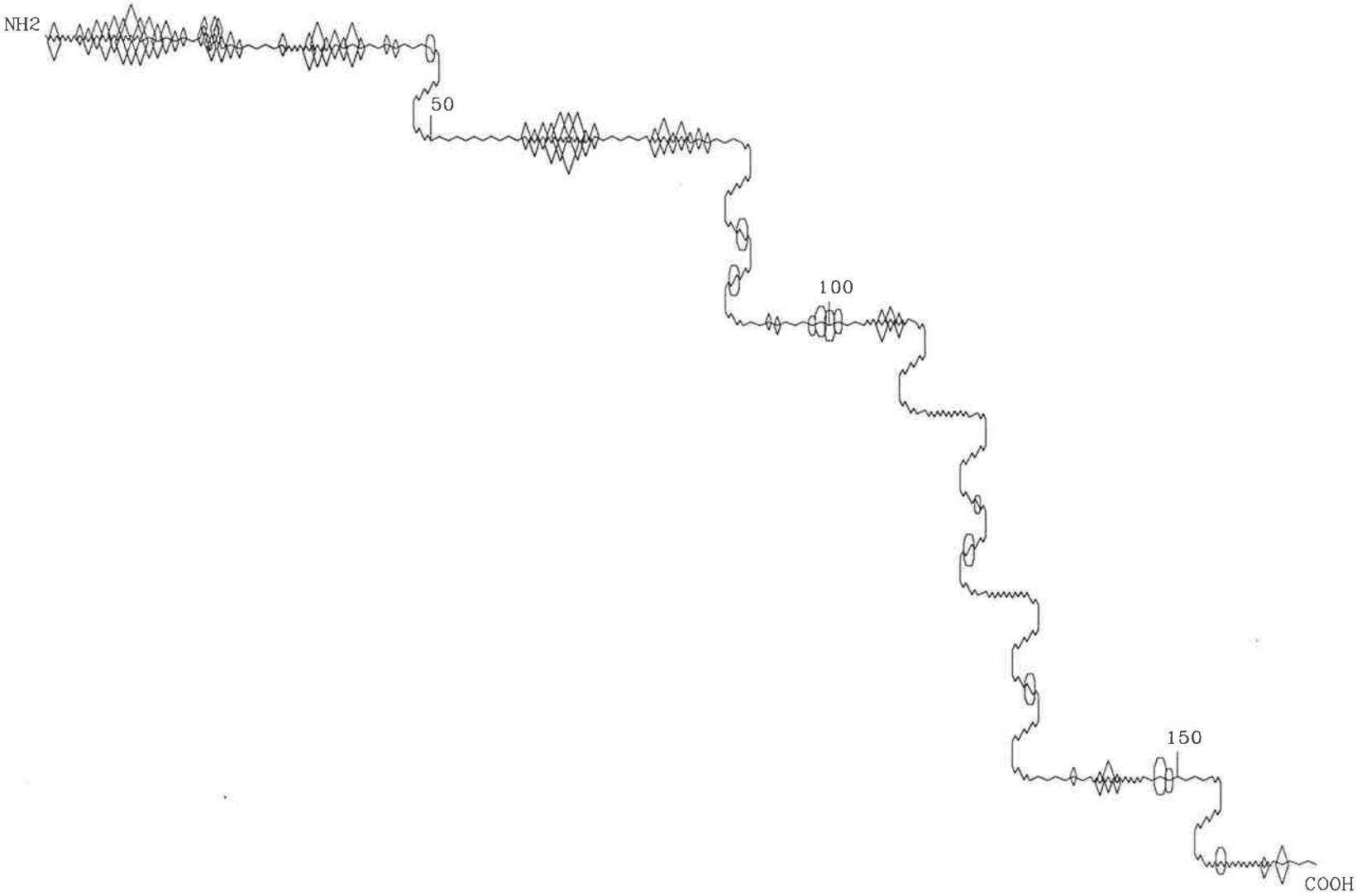

NH2
50
100
150
HOOC

Figure 4.16 Secondary structure and hydrophobicity profiles of CS3

The Chou and Fasman (1978) method for predicting the secondary structures was effected using the GCG programs, PEPTIDESTRUCTURE and PLOTSTRUCTURE. The hydrophobicity (◇) and hydrophilicity (○) predictions were by the Kyte and Doolittle (1982) method. A score of 1.3 or above for the hydrophobicity/hydrophilicity index was chosen for superimposing the latter properties on the secondary structure plot. Shown on the plot are, α-helix ( ), β-sheet ( ), β-turn ( ) and random coil ( ).


NH2
50
100
150
COOH

4.2.4.2 Cp82

The amino acid composition of Cp82 is given in Table. 4.1. Only four cysteine residues are present in this protein. It does not have an unusual number of any other amino acids. Although 11 Met residues were present, *in vivo* translation using ^{35}S-Met consistently produced a very faint band, which suggested that this protein was either translated at a low level or that it was rapidly turned over. The *in vitro* translation did not show its product. This probably reflects an inability of the *in vitro* system to translate long polypeptides, either due to the instability of mRNA or the depletion of amino acids.

The secondary structure predictions for the Cp82 polypeptide are summarised in Figure 4.10. Consistent with the absence of a predictable signal peptide cleavage site, no marked hydrophobic segment was found in the N-terminal region. Interestingly, all major hydrophobic domains on this protein occur in the domains shared by the ORFs contained within its coding region.

4.2.4.3 Cp61, Cp46, Cp31, Cp22 and Cp9

The amino acid compositions of all are given in Table 4.1 and the secondary structure prediction in Figures 4.11 to 4.15. Since these proteins, except Cp22, are fully contained within Cp82, they share the secondary structure profiles with the latter. One apparently significant finding was the absence of a hydrophobic region at the N-terminal end of Cp61, whereas hydrophobic domains were found in the N-terminal regions of the other four proteins.

4.2.4.4 CS3

The N-terminal region was markedly hydrophobic, characteristic of the leader peptide. No cysteine residue is present in this polypeptide. The secondary structure predictions are summarised in Figure 4.16, and Table 4.1 gives the amino acid composition.

4.2.5 Leader peptides and membrane spanning regions of the Cp proteins.

Examination of the amino acid sequences revealed signal peptides, according to the -1, -3 rule of von Heijne (1983), only for Cp27, Cp22 and CS3. The remaining proteins, Cp82, Cp61, Cp46, Cp22 and Cp9 did not reveal any convincing leader peptide. The latter could therefore be located in the cytoplasmic membrane or present within the cytoplasm itself. An attempt to decide between these two locations was made by using their hydrophobicity profile in conjunction with the predicted secondary structures.

A sequence of hydrophobic amino acid residues is an energetically favourable structure for a membrane spanning α-helix (Jennings, 1989). The length required to span a lipid bilayer plane, is 20 amino acid residues in an α-helix formation, however, this region may be smaller (10-12 aa) if the amino acids are in β-sheet formation. Examination of secondary structure profiles revealed that Cp82, Cp61 and Cp46 all had several hydrophobic domains in areas of β-sheet conformations (Figures 4.10, 4.11 and 4.12). The length of uninterrupted β-sheets varied from 5 and 11 residues. The secondary structure of Cp22 did not predict many such areas, with the only significant area being at the N-terminal region, and is probably in its leader peptide (Figure 4.15). Considering that these five proteins, Cp82, Cp61, Cp46, Cp22 and Cp9, do not have leader peptides, it is tempting to speculate that they are not transported to the outer membrane. They may either be located in the cytoplasmic membrane, or transported at lower efficiencies. This interpretation, however, apparently contradicts the possibility that these proteins may play a structural role in pilus synthesis and therefore must be located in the outer membrane (*see* section 4.2.7).

Although short stretches of hydrophobic domains are seen in Cp27, these are interrupted by strong hydrophilic regions (Figure 4.9). Presence of a leader peptide, conforming well with the rules, makes this a likely protein to be transported across the inner membrane. The absence of potential membrane spanning regions on this protein probably indicates that it is a periplasmic protein.

4.2.6 Potential roles of proteins.

It is evident from the N-terminal sequence determination of pili that *cpCS3* encodes the structural subunit. Mutational analyses indicated that Cp27 and Cp82 were directly involved in the biosynthesis of pili. Although an attempted mutational analysis of Cp22 was unsuccessful, a direct involvement of this protein could be inferred from the following observations:- (1) it shows homology to CS3 and PapH and, (2) pilus biogenesis apparently requires the suppression of an inframe UAG_{3578} codon in the coding region of Cp22. Involvement of other proteins in pili synthesis could not be ascertained because mutations that altered their products also affected Cp82. A sequence similarity search was done against the GenBank, NBRF-PIR and SWISS-PROT databases, in order to locate other ETEC proteins showing similarity to the Cp proteins and, thereby, gain an insight into the potential roles of these genes.

4.2.7 Protein similarities with other pilus biosynthesis genes.

Significant homologies to some proteins associated with Pap, K88 or K99 pilus biosynthesis were found, using the FASTA program of Pearson and Lipman (1988). These proteins were retrieved and individually compared to the Cp protein, using SEQDP, a program that calculates the statistical significance of homologies. Pairs of sequences, that showed a match of significance above 2 standard deviations, were further analysed by CLUSTAL (Higgins and Sharp, 1988), a sequence alignment program. Figures 4.17 to 4.19 show the multiple alignments of the sequences.

Cp27 showed a very significant match against the PapD (Lindberg *et al.*, 1989), FaeE and FanE (Bakker *et al.*, 1991) proteins of ETEC (Figure 4.17). These proteins function as the periplasmic transporters of the pilin subunits in the Pap, K88 and K99 pili systems. Its promoter region showed weak homology to that of the *cfaA* gene (Ghosal *et al.*, 1979; Hamers *et al.*,1989; Figure 3.7b, section 3.2.6.1b).

Cp82 showed significant homology (Figure 4.18) to the PapC (Norgren *et al.*, 1987), FaeD (Mooi *et al.*, 1986; Van Doorn *et al.*, 1982), FanD (Roosendaal and De

Figure 4.17 CLUSTAL alignment of Cp27 and other pili biosynthesis proteins

The amino acid sequences of Cp27, PapD, FanE and FaeE proteins were aligned by the program CLUSTAL (Higgins and Sharp, 1988). Residues that are identical in all sequences are indicated by a solid dot. Those residues which are conserved in all, but not necessarily identical, are indicated by open circles. Sequence numbers refer to the last residue in each line.


Cp27 MT---PIKLIFA------ALSLFPCSNIYANNITTQKFEAILGATRVIYHLDGNGESLRVK 52
FaeE MSKRNAVTTFFTNRVTKALGMTLALMMTCQS-AMASLA---VDQTRYIFRGDKDALTITVT 57
FanE MNK-------FIS----------IIALCVFSSY-ANAAFT---LNSTRYIYNEGQQSVSVNI- 41
PapD MIRKKILMA----------AIPLFV----------ISGADAAVSLDRTRAVFDGSEKSMTLDIS 44
Cp27 NPQISPILIQSKVMDEGSKDNAD----FIVTPPLFRLDAKRETDIRIVMVNGL---YPKDRE 107
FaeE NNDKERTFGGQAWVDNIVEKDTRPT-FVVTPSFFKVKPNGQQTLRIIMASDH---LPKDKE 114
FanE HNESEHKYGGQVWIDNIDKNGEVVF-FSPSPMVFKLNPKQKQIVRIVNINDN---LPKDRE 98
PapD NDNKQLPYLAQAWIENENQEKIITGPVIATPPVQRLEPGAKSMVRLSTTPDISKLPQDRE 104
Cp27 SLKTLCVRGIPPKQGDLWANNEKEFVGMKLNVSINTCIKLILRPHNLPKLDINSEGQIEW 167
FaeE SVYWLNLQDIPPALEG----------SG---IAVALRTKLKLFYRPKALLEGRKGAEEGISL 163
FanE SIFWLNVQEIPPAPKG----------DGGSLSLAINNRVKLIYRPIALKNGRDEAENNIKL 149
PapD SLFYFNLREIPPRSE----------KANVLQIALQTKIKLFYRPAAIKTRPNEVWQDQLI 154
Cp27 GIR-DGNLVAKNKTPYYFTIVNA-SFNGKALKTPGT--------LGPYEQ----------K 208
FaeE QSRPDGRTMLVNTTPYIFAIGSLLDGNGKKIATDNGTTQKLLMFMP-----GDEVQVKGN 218
FanE INSGTDSC-LENTTPYYFAISDV-KINGKSIDLNSDAKNKMGVFSPFSKVCLGNVNTSGN 207
PapD LNKVSGGYRIENPTPYYVTVIGL-----GGSEKQAEEGEFETVMLSPRSEQTVKSANYNTP 210
Cp27 LYTLPSKISVSGLVKWEIIGDLGESSETKKFNI-------- 241
FaeE VVKVDSLNDYGELQTWTINKKKPAAPEAAKAEKADTAEQK 258
FanE IT-VTAFNDYGVATSYTVQRSK---------- 228
PapD Y---LSYINDYGGRPVLSFICNGSRCSVKK----------EK 239

Figure 4.18 CLUSTAL alignment of Cp82 and other pili biosynthesis proteins

The amino acid sequences of Cp82, PapC, FimD, FanD and FaeD proteins were aligned by the program CLUSTAL (Higgins and Sharp, 1988). Residues that are identical in all sequences are indicated by a solid dot. Those residues which are conserved in all, but not necessarily identical, are indicated by open circles. Sequence numbers refer to the last residue in each line.

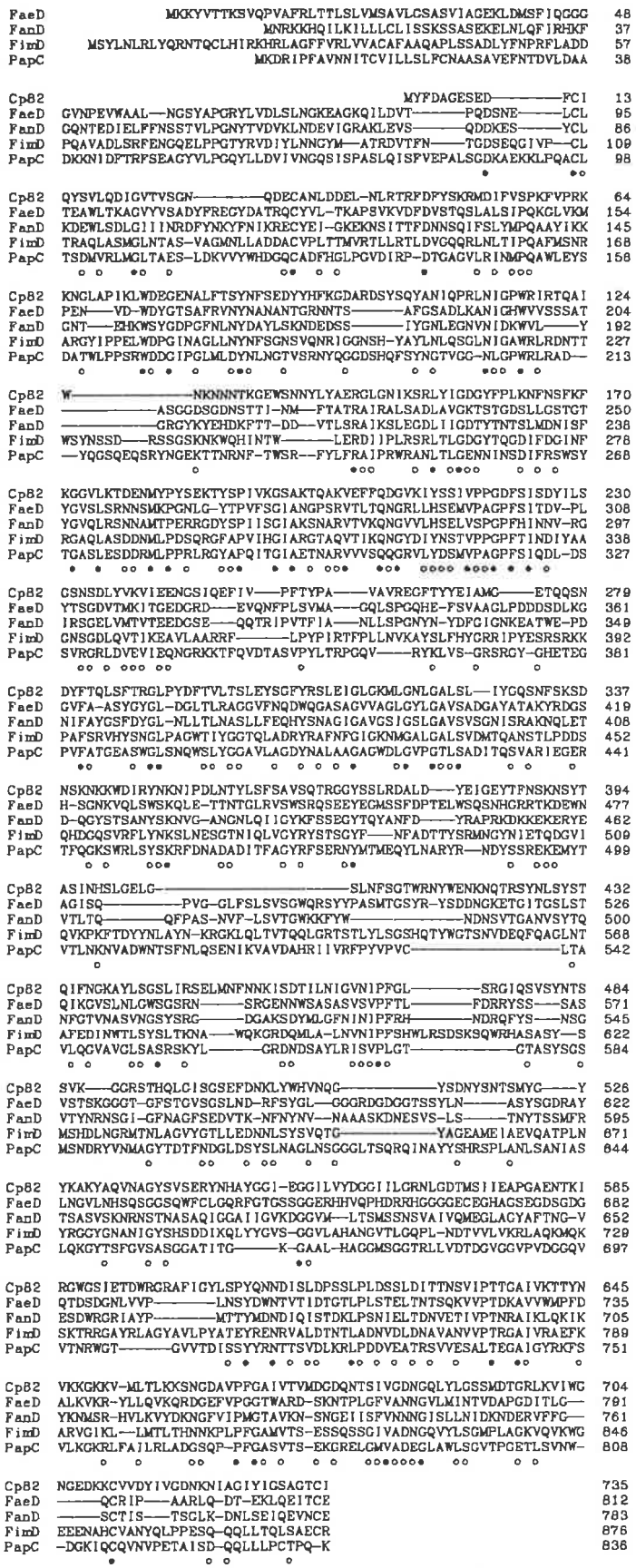


Figure 4.19 Homology of Cp22 and CS3 to each other and to other ETEC proteins

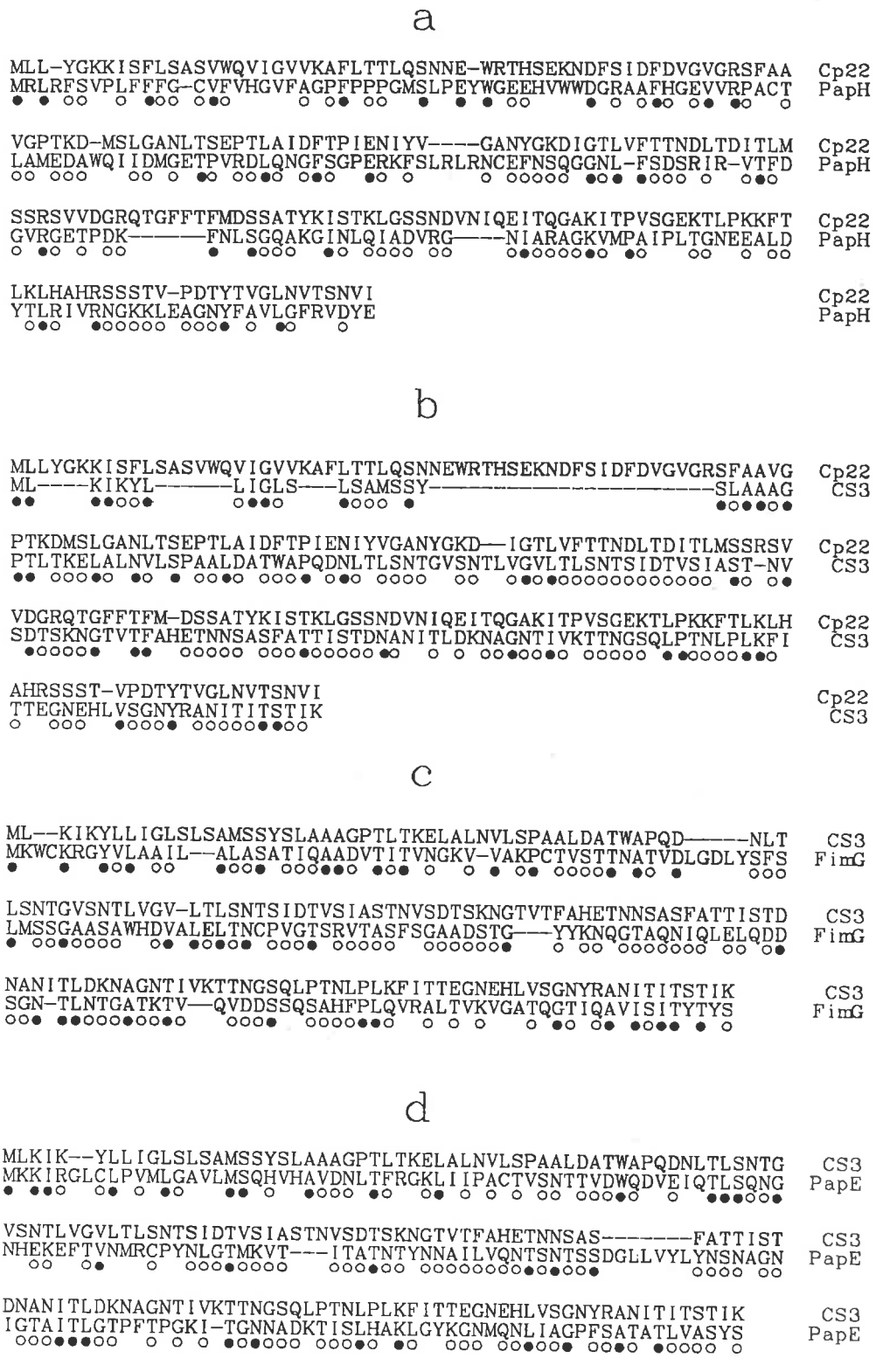
Comparisons of the amino acid sequences. The sequences were individually aligned by the program CLUSTAL (Higgins and Sharp, 1988). Solid circles indicate identical residues whereas open circles mark the conserved replacements.

a - Cp22 vs. PapH

b - Cp22 vs. CS3

c - CS3 vs. FimG

d - CS3 vs. PapE


a
MLL-YGKKISFLSASVWQVIGVVKAFLTTLQSNNE-WRTHSEKNDFSIDFDVGVGRSFAA Cp22
MRLRFSVPLFFFG-CVFVHGVFAGPFPPPGMSLPEYWGEEHVWWDGRAAFHGEVVRPACT PapH
VGPTKD-MSLGANLTSEPTLAIDFTPIENIYV----GANYGKDIGTLVFTTNDLTDITLM Cp22
LAMEDAWQIIDMGETPVRDLQNGFSGPERKFSLRLRNCEFNSQGGNL-FSDSRIR-VTFD PapH
SSRSVVDGRQTGFFTFMDSSATYKISTKLGSSNDVNIQEITQGAKITPVSGEKTLPKKFT Cp22
GVRGETPDK------FNLSGQAKGINLQIADVRG----NIARAGKVMPAIPLTGNEEALD PapH
LKLHAHRSSSTV-PDTYTVGLNVTSNVI Cp22
YTLRIVRNGKKLEAGNYFAVLGFRVDYE PapH
b
MLLYGKKISFLSASVWQVIGVVKAFLTTLQSNNEWRTHSEKNDFSIDFDVGVGRSFAAVG Cp22
ML-----KIKYL------LIGLS---LSAMSSY----------SLAAAG CS3
PTKDMSLGANLTSEPTLAIDFTPIENIYVGANYGKD—IGTLVFTTNDLTDITLMSSRSV Cp22
PTLTKELALNVLSPAALDATWAPQDNLTLSNTGVSNTLVGVLTLSNTSIDTVSIAST-NV CS3
VDGRQTGFFTFM-DSSATYKISTKLGSSNDVNIQEITQGAKITPVSGEKTLPKKFTLKLH Cp22
SDTSKNGTVTFAHETNNSASFATTISTDNANITLDKNAGNTIVKTTNGSQLPTNLPLKFI CS3
AHRSSST-VPDTYTVGLNVTSNVI Cp22
TTEGNEHLVSGNYRANITITSTIK CS3
c
ML--KIKYLLIGLSLSAMSSYSLAAAGPTLTKELALNVLSPAALDATWAPQD-----NLT CS3
MKWCKRGYVLAAIL--ALASATIQAADVTITVNGKV-VAKPCTVSTTNATVDLGDLYSFS FimG
LSNTGVSNTLVGV-LTLSNTSIDTVSIASTNVSDTSKNGTVTFAHETNNSASFATTISTD CS3
LMSSGAASAWHDVALELTNCPVGTSRVTASFSGAADSTG---YYKNQGTAQNIQLELQDD FimG
NANITLDKNAGNTIVKTTNGSQLPTNLPLKFITTEGNEHLVSGNYRANITITSTIK CS3
SGN-TLNTGATKTV—QVDDSSQSAHFPLQVRALTVKVGATQGTIQAVISITYTYS FimG
d
MLKIK--YLLIGLSLSAMSSYSLAAAGPTLTKELALNVLSPAALDATWAPQDNLTLSNTG CS3
MKKIRGLCLPVMLGAVLMSQHVHAVDNLTFRGKLIIPACTVSNTTVDWQDVEIQTLSQNG PapE
VSNTLVGVLTLSNTSIDTVSIASTNVSDTSKNGTVTFAHETNNSAS-------FATTIST CS3
NHEKEFTVNMRCPYNLGTMKVT---ITATNTYNNAILVQNTSNTSSDGLLVYLYNSNAGN PapE
DNANITLDKNAGNTIVKTTNGSQLPTNLPLKFITTEGNEHLVSGNYRANITITSTIK CS3
IGTAITLGTPFTPGKI-TGNNADKTISLHAKLGYKGNMQNLIAGPFSATATLVASYS PapE

Graaf, 1989) and FimD (Klemm and Christiansen, 1990) of *E. coli* .

A minor Pap fimbrial component is encoded by PapH, in the Pap pilus system. A unique role for the PapH product in anchoring the fimbria to the cell membrane has been proposed. Comparison of Cp22 against this protein showed sufficient homology to suggest that Cp22 probably played a similar role (Figure 4.19 a). More significant homology was found between Cp22 and the CS3 subunit (Figure 4.19 b).

CS3 showed weak homology to FimG and PapE proteins (Figure 4.19 c and d). The latter two function as minor pilin components (Hanson and Brinton, 1988; Lindberg *et al.*, 1987). No significant homology between CS3 and any of the other pilin subunit proteins was found.

4.2.8 Intracellular localisation of the gene products

An investigation of the intracellular location of the various gene products may be helpful in assigning them possible roles in pilus biosynthesis. This was done using ^{35}S-labelled minicells, which were fractionated into three components, as periplasm, inner membrane and outer membrane. (Cytoplasmic fraction was also taken, but it did not show any difference from the whole-cell sample. Presumably, it was contaminated either by whole cells or by the inner/outer membrane fractions). The protein products were visualised by autoradiography after separating on SDS-PAGE (Figure 4.20). Although every attempt was made to separate the two membrane fractions, cross contamination could not be ruled out. For this reason, although the two lanes are separately labelled in Figure 4.20, a distinction between the two fractions is not made in the following discussion.

The mature form of Cp27 was found almost entirely in the periplasm (Figure 4.20, lane *d*). The membrane fractions showed a band corresponding to its precursor form. The mature form of CS3 too was found in the periplasm (Figure 4.20, lane *d*). Contrary to the expectation, only a faint band corresponding to the CS3 subunit was found in the membrane fractions. Probably, the forces exerted during minicell purification

Figure 4.20 Intra-cellular location of proteins

Minicells carrying pBR322 and pPM484 were labelled with ^{35}S-Met and then fractionated into periplasmic, outer membrane and inner membrane components. Lanes *a* and *b* represent the whole translation products of minicells. The entire amount of each fraction (fractionated from 250 ml of overnight cultures) were run in SDS on 15% PAGE and visualized by autoradiopgraphy.

Lanes:

a - Minicell lysate carrying pBR322

b - Minicell lysate carrying pPM484

c - Periplasmic fraction (pBR322)

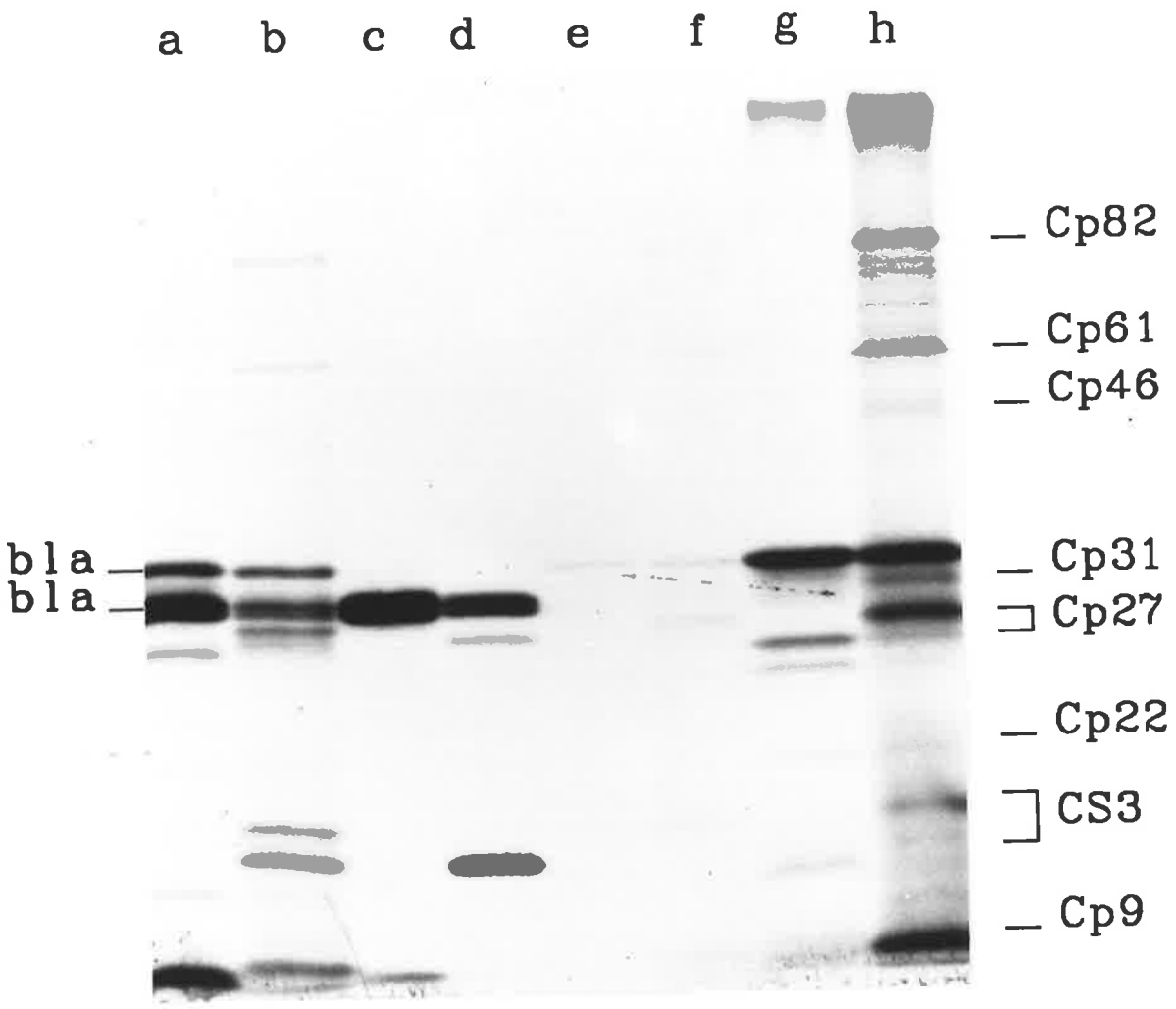
d - Periplasmic fraction (pPM484)

e - Outer membrane fraction (pBR322)

f - Outer membrane fraction (pPM484)

g - Inner membrane fraction (pBR322)

h - Inner membrane fraction (pPM484)


a
b
c
d
e
f
g
h
bla
bla
Cp82
Cp61
Cp46
Cp31
Cp27
Cp22
CS3
Cp9

and the fractionation procedures would have sheared the CS3 pili from the cell surface, thus explaining the failure to detect it in large amounts in the membrane fractions.

Five proteins, Cp82, Cp61, Cp46, Cp31 and Cp9 were found in the membrane fractions alone (Figure 4.20, lanes *f* and *h*). Two closely spaced bands were noted for Cp82 which were similar in appearance to the bands visible in the whole cell fraction (Figure 4.20, lanes *b*). A very faint band corresponding to Cp22 was found in the membrane fraction (Figure 4.20, lane *h*).

4.3 Discussion

4.3.1 Analysis of translational initiation

The first objective of the work described in this chapter was to investigate whether the five overlapping polypeptides originated by independent initiations of translations or by post-translational processing of a larger, presumably that of Cp82, polypeptide. In this regard, evidence for independent translation was obtained as follows:-

(1) the *Nco*I mutation that resulted in the disappearance of Cp82 retained the other polypeptides, proving that they did not originate from the former,

(2) the *Pst*I frameshift within Cp61 retained the polypeptides of Cp46, Cp31 and Cp9 and, therefore, these could not have originated from the 61 kDa polypeptide, and,

(3) the *Bgl*II deletion resulted in the disappearance of the 46 kDa band but not of the 31 and 9 kDa bands. The latter two proteins cannot, therefore, be the processed forms of the 46 kDa polypeptide.

It is, however, possible that the 9 kDa protein is derived from the 31 kDa protein. This seems unlikely, as there is a candidate initiation codon, with RBS, for its translation at an appropriate distance from the termination codon. The CS3 gene cluster, therefore, is unique among the pilus biosynthesis operons by having independent translational units in an overlapping arrangement.

Although the same reading frame is shared by all five ORFs, this does not

necessarily mean that they are translated from the same mRNA. For instance, it is possible that five overlapping transcripts are produced from separate promoters, and each is translated to give a single polypeptide. Such a situation may be necessary to avoid problems encountered with the entry of fresh ribosomal subunits into the mRNA that is being actively translated and, thereby, covered with ribosomes. Alternatively, one or more mRNAs may be translated from different internal initiation codons at different times. Examination of the pattern of transcription, which forms the subject of next chapter, may resolve those possibilities.

4.3.2 Mutational analyses to identify potential roles

The indispensable nature of the various genes in the expression of pili on the cell surface was also investigated. It was found that the *Eco*RI mutation in *cp27* and the *Nco*I mutation in *cp82* resulted in the absence of surface-pili. Products of these two genes are therefore directly involved in pilus biogenesis. Mutagenesis of the remaining genes also resulted in the absence of pili, but, given the concomitant effect on Cp82 of any mutation in the overlapping regions, the absence of pilus synthesis in these cases could not be considered as evidence for their direct involvement. However, the products of *cp61*, *cp46*, *cp31* and *cp9*, like that of *cp*82, are all associated with the membrane. Thus, although no direct involvement could be demonstrated for these proteins, it could be assumed that they played some structural role in pilus biogenesis. Homology between Cp27 and the periplasmic transporter/chaperone proteins from other pili systems (e.g. PapD of Pap pilus system; Lindberg *et al.*,1989; Kuehn *et al.*, 1991) indicates that this protein probably functions in a similar manner in CS3 pilus biogenesis. Almost all of the mature form of this protein was found in the periplasmic fraction, which supports the above conclusion. A possible role of this protein could be in the folding and/or transport of the CS3 subunit prior to assembly into cell surface pili. In this regard it would probably function catalytically, requiring it to be produced only in small amounts. However, its band intensity during minicell labelling suggests that it is produced at

significantly higher level.

Judging from its homology to the channel proteins in other pilus systems, the product of Cp82 might be a channel protein as well. This protein, however, is either produced in very small amounts or is being used up very rapidly, as its band in minicell translations is always very faint.

The proteins, Cp46, Cp31 and Cp9 are also seen in the membrane fractions and, like Cp82 and Cp61, these too may perform some structural roles.

Cp22 showed homology to CS3 and also to PapH (Båga *et al.*, 1987), suggesting that it is probably a basal protein, involved in the last step of the assembly of CS3 pili. PapH has been proposed to function as the length modulator, by incorporating into the base of the pap pili and terminating further elongation. A similar role of Cp22 in CS3 pilus biogenesis may be predicted.

4.3.3 Identification of potential roles by cell fractionation

The fractionation of minicells implied that the 82, 61, 46, 31 and 9 kDa proteins were located in the membranes. A distinction between the inner and outer membrane fractions could not be made, as cross contamination could have occurred, and minicell fractionations are notoriously fraught with problems (Achtman *et al.*, 1979). The presence of the precursor forms of the 27 kDa and CS3 polypeptides in the membrane fractions suggests that they probably represent the molecules that are trapped as they pass through the inner membrane.

Two species of the 22 kDa protein have been observed in both *in vivo* and *in vitro* translations. Since processing does not occur in *in vitro* translations, these two forms could have resulted from independent initiations or by a mechanism of cleavage other than that mediated by signal peptidase. The major, shorter species is seen in the membrane fractions. This is consistent with the earlier assumption that it might form part of the pilus. Minicell translation of the *Bgl*II deletion mutant showed an accumulation of this shorter band. As pili are not formed in this mutant it probably suggests that Cp22 is

accumulated in either the periplasm or cytoplasm. Interestingly, in the same sample that showed an accumulation of Cp22, the precursor form of CS3 subunit is in far excess over the mature form (Figure 4.3A, lane *j*). However, no accumulation of either Cp22 or CS3 precursor was seen when pilus biogenesis was inhibited by mutating the *cp27* and *cp82* ORFs.

Though surface pilus synthesis was absent, none of the mutant plasmids showed the disappearance of the CS3 polypeptide in minicell translations. This is an indication that these gene products do not take part in the regulation of expression of CS3. Examination of the DNA sequence reveals a potential promoter for CS3. The RBS for this gene is also highly homologous to the Shine-Dalgarno sequence. It was, therefore, not surprising to see that absence of any of the other proteins did not result in a loss of CS3 subunit polypeptide.

An alternative processing site for the CS3 precursor polypeptide has been identified. Processing apparently occurs at both these sites in the wild type ETEC strains as indicated by the appearance of two bands in crude pili preparations from PB176. Only one band, corresponding to the minor form, has been observed in *E. coli* K-12. This suggests that only one processing site, the one after the 14^{th} residue, is able to be recognized in *E. coli* K-12. Since CS3 pili were present in *E. coli* K-12 that harboured pPM484, this suggests that an extra 7 aa at the N-terminal end do not affect its function, or ability to be assembled.

Examination of the secondary structure profiles of the Cp proteins indicates that all are devoid of extensive α-helices, but rich in β-sheets. Marked hydrophobic regions occur along the length of Cp82. Significantly, most of these occur near the predicted start sites of the smaller polypeptides contained within it. Several of the predicted hydrophobic domains overlapped these β-sheets, which probably signifies that they are the membrane spanning regions of the protein molecules.

In conclusion, results presented in this chapter show that all polypeptides

identified by minicell translations originate by independent initiation of translations. They all seem to take part in CS3 pilus biogenesis, although direct involvement of only Cp27, Cp82 and CS3 has been demonstrated. Future work using monoclonal antibodies and immunogold EM studies may indicate specific roles of each of these proteins.

CHAPTER 5

Transcriptional Analysis Of The CS3 Operon

CHAPTER 5

Transcriptional Analysis Of The CS3 Operon.

5.1 Introduction

The control of gene expression in bacteria occurs mainly at the level of transcription (Rosenberg and Court, 1979). Genes involved in a common biosynthetic pathway are often grouped together in transcriptional units called operons. Even though the arrangement of the *cp* genes suggests that they form an operon, sequence analysis implied the existence of several putative promoters. The most likely candidates are marked on the sequence at nt 341, 1797, 2574 and 4135. These are designated P_{27}, P_{61}, P_{31} and P_{CS3}, respectively (Figure 3.4a and 3.6). No putative promoter for *cp82* was detected, nor was there a Rho-independent terminator between the end of *cp27* and the start of *cp82*. Other putative promoters detected by computer analysis could not all be disregarded, even though most of these are likely to have been spurious and were erroneously detected due to the high A+T content of the DNA.

In this chapter, the number, size and location of transcripts from the CS3 region and the relative strengths of the promoters are investigated. The emphasis of these studies centres on identifying the major transcripts that encode the genes necessary for pilus synthesis, and characterising the putative promoters, P_{27} and P_{CS3}. Other putative promoters within the coding region of *cp82* have not been investigated experimentally, apart from showing that transcription does initiate in the vicinity of some of these promoters.

5.2 Results

5.2.1 Growth conditions for studying the transcription patterns

The expression of pili is affected by growth conditions such as temperature and composition of the growth medium (De Graaf *et al.*, 1980; Gaastra and De Graaf, 1982). This could mean that either the transcription of pilus genes is affected by the growth

conditions, or pilus biogenesis is prevented by some other physiological changes to the cells. In the case of Pap pili, transcription of the structural gene for the pilus subunit (*papA*) is regulated in response to the growth temperature so that pilus formation is absent at low temperatures (Göransson and Uhlin, 1984). Glucose and other fermentable sugars suppress the synthesis of type I (Eisenstein and Dodd, 1982) and K99 fimbriae (Isaacson, 1980b; Ollier and Girardeau, 1983). Indications have been that this glucose effect is a manifestation of selective outgrowth of non-piliated cells in the population rather than a transcriptional regulation of pilus synthesis (Van Verseveld *et al.*, 1985).

It has been observed, while preparing crude CS3 pili, that pili were much less abundant on cells grown in nutrient broth (NB) than on those grown in colonization factor broth (CFB) or Luria Broth (LB). In order to determine whether this difference reflected the effect of these media on transcription, cultures were grown in different media and total cellular RNA was prepared from them. An analysis of the dot blots, using a pPM484-specific probe, did not reveal any significant difference between the CFB and LB culture media. However, significantly lower levels of target mRNA were detected in the cells grown in NB (Figure 5.1 A). To examine the effect of temperature on pilus synthesis, cells were grown at 18°C and pilus synthesis examined, and found to be negative, by immunogold EM and colony blotting (data not shown). All subsequent experiments were therefore done with RNA prepared from cells grown, at 37°C, in either CFB or LB.

5.2.2 Northern blot analysis of transcripts

Computer analyses have predicted putative promoters on both strands of the DNA even though all putative ORFs are on the same strand. In order to establish whether transcription occurs in one or both directions, and to determine the number and sizes of transcripts, Northern blots of total RNA from cells carrying pPM484 were hybridised with single and double-stranded DNA probes. The single-stranded probes were prepared from M13 clones of restriction fragments, whereas double-stranded

Figure 5.1 Relative abundance of mRNA in cells grown in different culture media

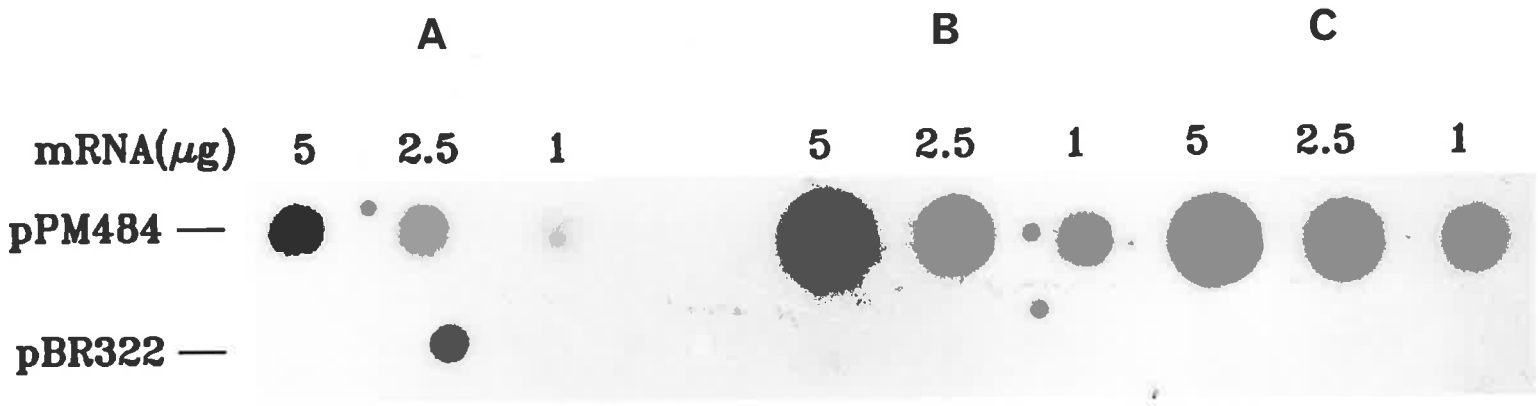
Dot blots of total RNA prepared from exponentially growing cultures of *E. coli* strains harbouring either pPM48 or pBR322 were probed with the 4.7 kb *Hin*dIII fragment from pPM484. Approximate amounts (μg) of mRNA loaded in each spot are marked.

Groups:

A - Nutrient broth.

B - L Broth.

C - CF Broth.


A
B
C
mRNA(μg)
5
2.5
1
5
2.5
1
5
2.5
1
pPM484 —
pBR322 —

probes were made from the entire insert DNA in pPM484 or from an internal *Sca*I fragment (Figure 5.2, probe 4). The RNA isolated from cells carrying pBR322 was used as control. It was found that at least four bands (4.3, 2.9, 2.0 and 0.6 kb) were detected (Figure 5.3, lane *a*) with the single-stranded probe (probe 1 in Figure 5.2), which is derived from the CS3 coding region and is complementary to the mRNA. The longest transcript (4.3 kb) probably originates in the vicinity of P_{27} and the shortest (0.6 kb) in the vicinity of P_{CS3}. One transcript of 2.9 kb could be mapped to the region around nt 1800 and was assumed to originate from P_{61} (Figure 5.3; lanes *a* and *h*). Another transcript of 2.0 kb probably corresponded to the mRNA initiating from P_{31}. The latter was confirmed by using the *Sca*I fragment (nt 1358 and 2788) as a double-stranded probe, which revealed the 2.0 and 2.9 kb bands after longer exposure (Figure 5.3; lanes *g* and *h*). This probe did not detect any other band. The 0.6 kb band, assigned to P_{CS3}, is downstream to this probe and would not be detected, whereas the 4.3 kb band might have been missed due to its relatively low level of expression.

No transcript was detected when the probes were from the opposite strand (probes 2 and 3 in Figure 5.2) or with pBR322 (Figure 5.3, lanes *b* to *d*). Thus, all transcripts encoded on the DNA in pPM484 have the same polarity as suggested by the potential ORFs. As expected, a double-stranded probe, prepared by using the entire insert from pPM484 revealed an identical pattern of bands (Figure 5.3, lane *f*) as that seen with the single-stranded probe. Since no new band was detected by this probe, as compared to those detected by the probe within the CS3 coding region, it indicated that all transcripts extended to the end of *cpCS3*. This finding confirmed the earlier prediction that no Rho-independent terminator was present upstream of the CS3 coding region.

Northern blots were also prepared with RNA isolated from the ETEC strain, PB176, in order to verify whether any difference in the pattern of transcription existed between ETEC and *E. coli* K-12. Upon probing with the entire insert from pPM484, two bands, identical in length to the P_{CS3} and P_{31} transcripts, were readily visible (Figure 5.3,

Figure 5.2 Linear diagram showing probes, putative promoters and their transcripts.

Thick horizontal lines with arrow heads on both ends are the double-stranded probes, whereas those with single arrow heads are M13 single-stranded probes. Probe numbers, referred to in the text, are indicated. Length and direction of transcripts are shown by thin lines with arrow heads. Fragments cloned in the pCB vectors (A to D) are indicated. These have been cloned in both possible orientations, so that in one orientation the promoter activity will be indicated on BCIG plates whereas in the other orientation the activity will be indicated on GalK plates. Arrow heads represent the putative promoters (P_{27}, P_{61}, P_{31} and P_{CS3}). Vertical bars indicate the sequence positions (kb).

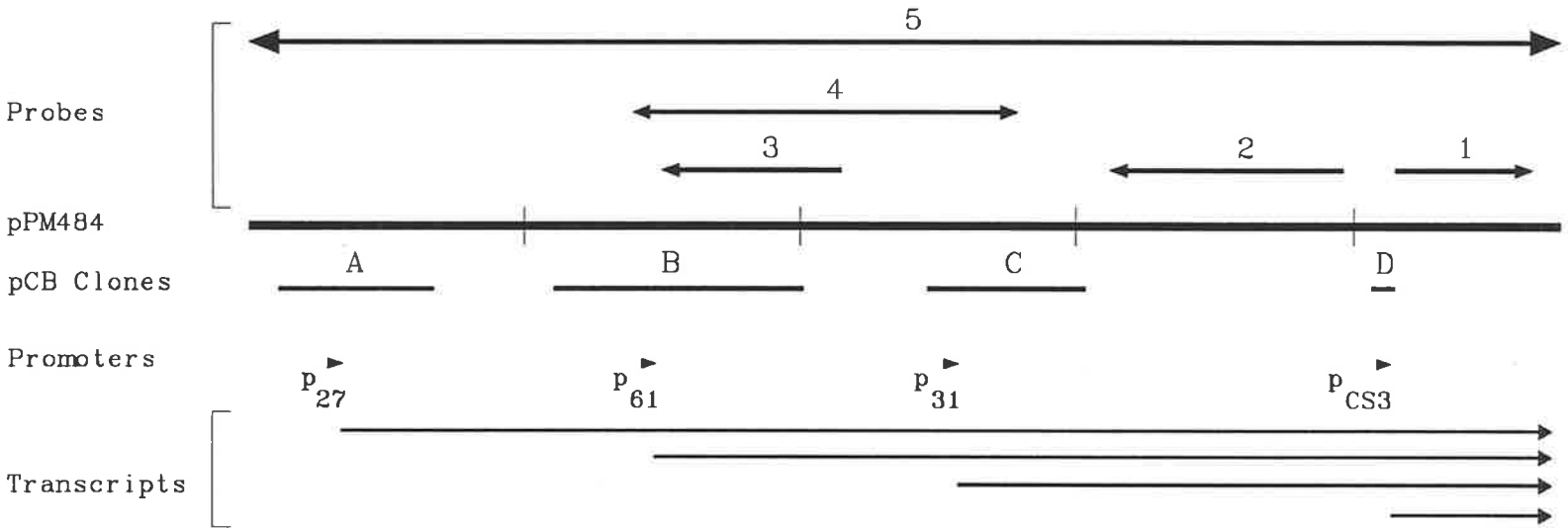

5
Probes
4
3
2
1
pPM484
A
B
C
D
pCB Clones
Promoters
p27
p61
p31
pCS3
Transcripts

Figure 5.3 Messenger RNA detected by pPM484-specific probes.

Approximately 10 μg each of total RNA, isolated from late exponential phase cultures of PB176 and *E. coli* K-12, carrying pPM484 or pBR322, were run on 1.2 % agarose, transferred to nitrocellulose and hybridized to probes (*see* Figure 5.2).

Panels:

A - Single-stranded probes

B - Double stranded probe of the 4.7 kb *Hin*dIII fragment

C - Double-stranded probe of the *Sca*I fragment (nt 1360-2788)

Positions of the bacterial 23S (3700 nt) and 16S (1700 nt) ribosomal RNA are marked. Approximate sizes of the bands are indicated.

Lanes (*see* Figure 5.2 for probe numbers):

a - probe 1; RNA from cells carrying pPM484

b - probe 1; RNA from cells carrying pBR322

c - probe 2; RNA from cells carrying pPM484

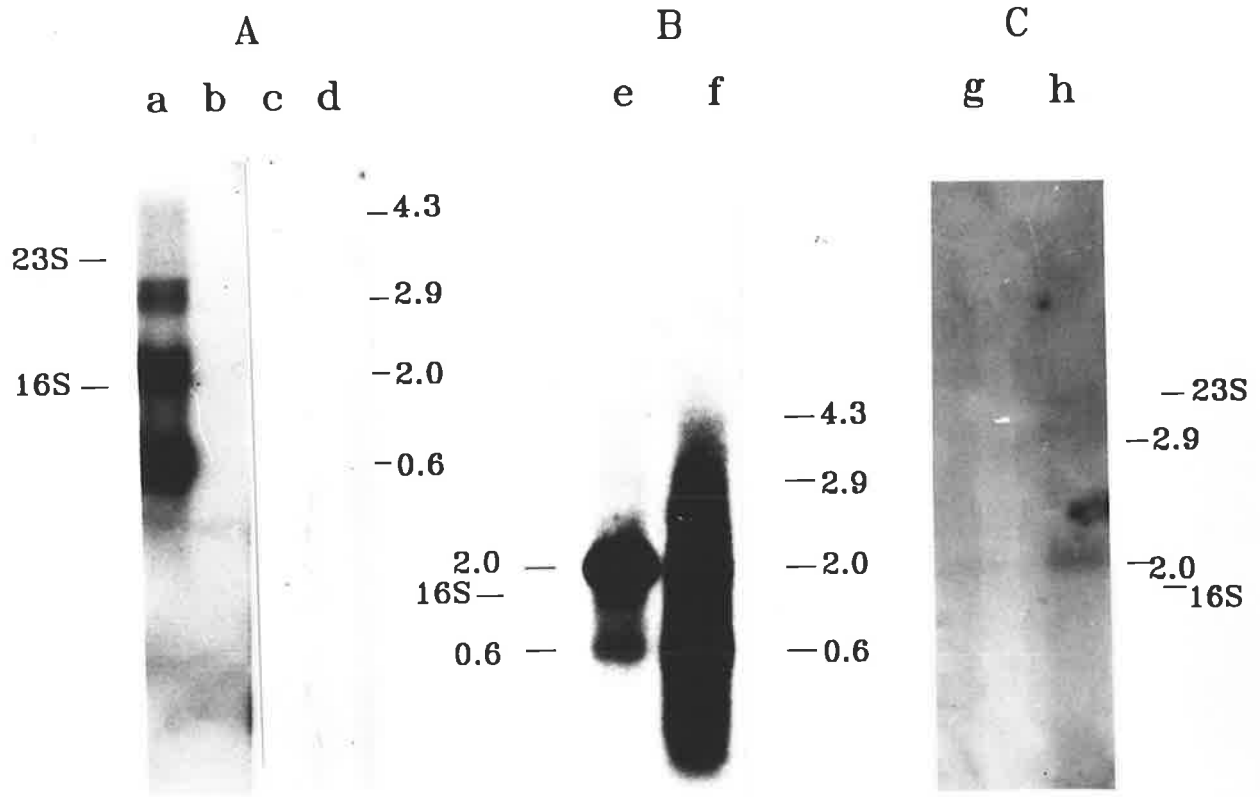
d - probe 3; RNA from cells carrying pPM484

e - probe 5; RNA from PB176

f - probe 5; RNA from cells carrying pPM484

g - probe 4; RNA from PB176

h - probe 4; RNA from cells carrying pPM484


A
a b c d
23S
16S
4.3
2.9
2.0
0.6
B
e f
2.0
16S
0.6
4.3
2.9
2.0
0.6
C
g h
23S
2.9
2.0
16S

lane *e*) and two faint bands, probably corresponding to the P_{27} and P_{61} transcripts appeared after longer exposure of the blot to X-ray film. Thus, the number and sizes of transcripts appear to be the same in PB176 as in *E. coli* K-12, indicating a similar pattern of transcription in both organisms. However, a difference in the relative abundance of mRNAs, originating from P_{31} and P_{CS3}, was observed between the two organisms. Thus, while the 600 nt mRNA band, originating from P_{CS3}, was much more intense than the 2.0 kb band in *E. coli* K-12, the reverse was true in the case of PB176 (Figure 5.3, lanes *a* and *e*).

The fact that transcripts from PB176 are identical in size to those from pPM484 suggests that termination of transcription occurs within the DNA cloned in pPM484. The stem-loop structure downstream of the *cpCS3* ORF (Figures 3.4a and 3.8) is therefore the putative terminator of all transcripts.

5.2.3 Localisation of P_{27} and P_{CS3} by primer extension

Primer extensions were performed in order to map the 5' ends of P_{27} and P_{CS3} transcripts. An oligonucleotide primer, complementary to nucleotides 399 to 420, should produce an approximately 63 nt extension product from the P_{27} transcript, whereas one complementary to nucleotides 4248 to 4272 should give a 131 nt product from the P_{CS3} transcript. Primer extensions on the total RNA from cells carrying pPM484 revealed several fragments, of which the major ones approximately corresponded in length to the above predicted sizes (Figures 5.4 and 5.5). No primer extension was done to confirm the putative promoters detected upstream to the *cp61* and *cp31* ORFs.

5.2.4 *In vivo* analyses of promoter activities

The promoter-detection vectors, pCB182 and pCB192 (Schneider and Beck, 1986) could be used in the *in vivo* detection of promoter activities on cloned DNA fragments. Depending on the orientation of the insert, an active promoter would express either the β-galactosidase (*lacZ*) or galactokinase (*galK*) gene. In order to confirm the *in vivo* activity of the four promoters, predicted by the Northern blot, *Sau*3AI restriction

Figure 5.4 Primer extension on P27 transcript

Total RNA isolated from cells harbouring pPM484 was used in primer extension with an oligonucleotide complementary to nt 399 to 420. Two major bands, marked as P_{27} and $P_{27'}$, correspond to mRNA initiations from nt 281 and 348. The sequencing ladder shown here was generated by dideoxy sequencing of double-stranded DNA by using the same oligonucleotide as primer. The nucleotide positions of bands are indicated by arrows.

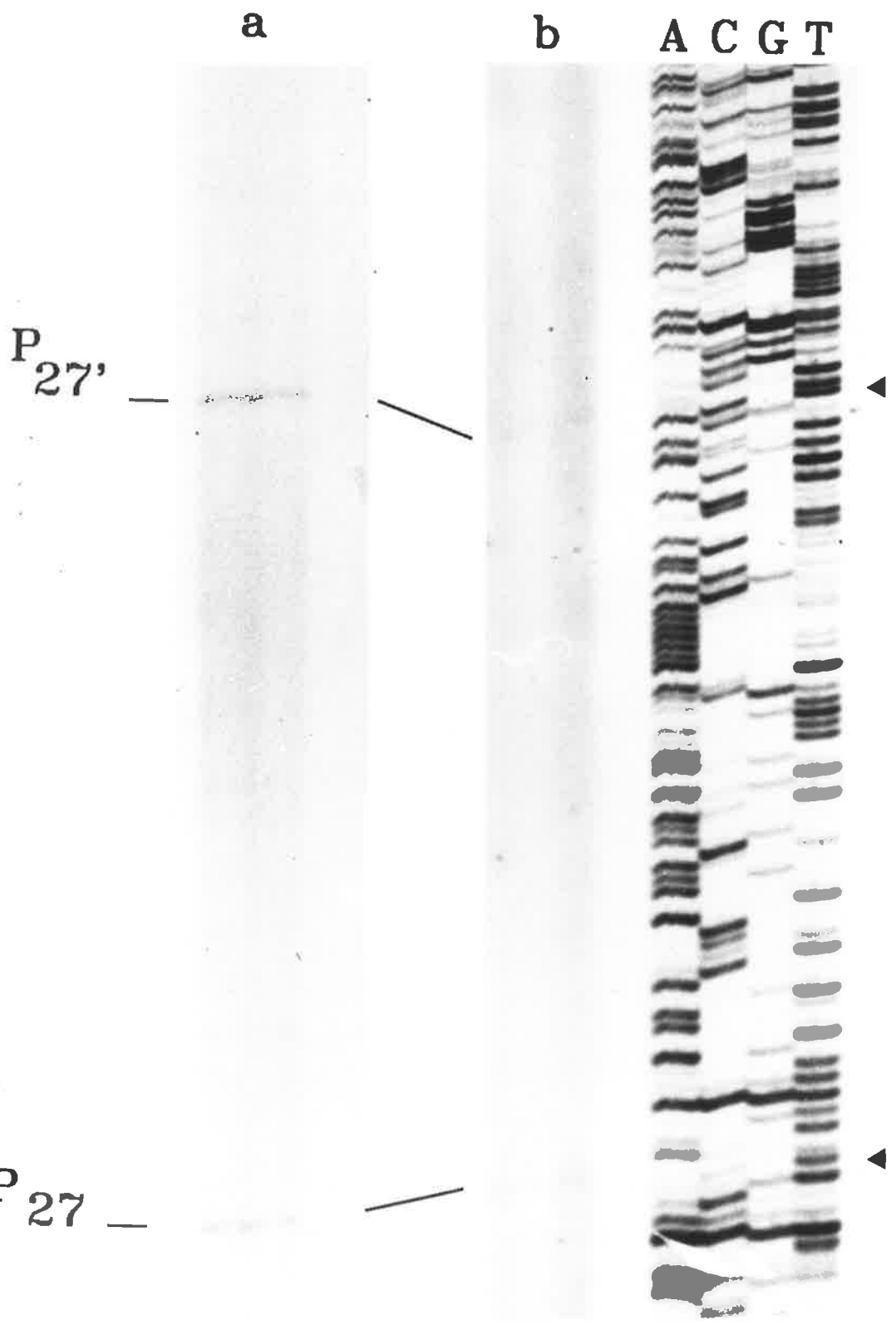
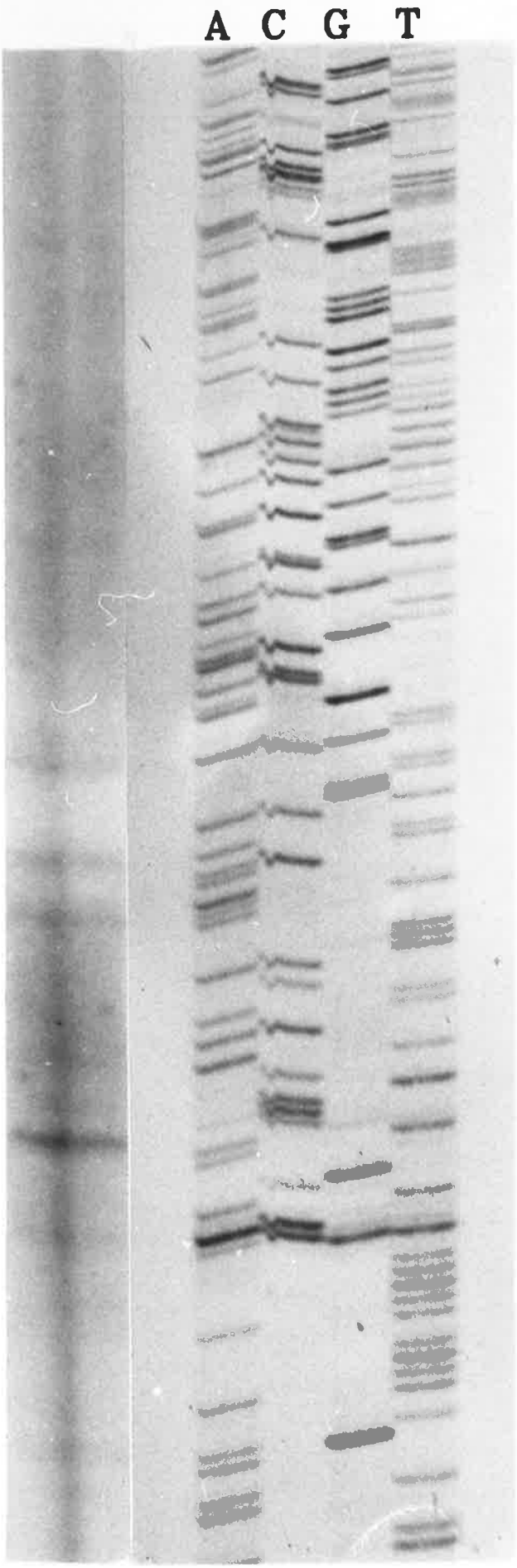

a
b
A C G T
P27'
P27

Figure 5.5 Primer extension on PCS3 transcript

Total RNA isolated from cells harbouring pPM484 was used in primer extension with an oligonucleotide complementary to nt 4248 to 4272. The sequencing ladder generated by the same primer is used as size marker. The major band, marked by arrow, corresponds in size to mRNA initiation from nt 4149.


A C G T

fragments encoding P_{61}, P_{27}, P_{31} and P_{CS3} were cloned in pCB182/pCB192 and the colonies tested on appropriate indicator plates. The results, summarised in Figure 5.6 and Table 5.1, indicate that detectable promoter activities are present on fragments encoding P_{27} and P_{CS3}, when the cloned promoter fragments were in either direction (pCB clones of fragments A and D in Figure 5.2). The fragments encoding P_{61} and P_{31} (B and C in Figure 5.2) showed β-galactosidase activity when their direction of transcription was the same as that of *lacZ*, but no color reaction on the GalK plate could be observed when transcription was in the direction of the *galK* gene. However, low level of activity for P_{31}, but not for P_{61}, was seen when the GalK assay was performed. The fragment containing P_{CS3} had been predicted to carry another promoter on the complementary strand. This was found to be active, albeit at a lower level than P_{CS3}.

5.2.5 Estimation of relative promoter strengths by GalK assay

Northern blotting revealed that the transcripts from P_{CS3} and P_{31} were more intense than that from P_{27}. Differences in mRNA levels could result from differences in the relative promoter strengths or mRNA stabilities (Hoopes and McClure, 1987). The P_{27} transcript is longer than those originating from P_{CS3} and P_{31}, and this could result in decreased stability. Alternatively, its promoter could be less active *in vivo*. The pCB system can be used in the testing of promoter strengths by quantitating the GalK enzyme activity (McKenny *et al.*, 1981). These assay results are presented as mean values of three independent experiments in which the GalK activity was determined in duplicates (Table 5.1). In order to compare the relative strengths of the promoters, a control experiment was set up with the *lac* promoter (P_{lac}), which gave 620 GalK units. This correlates with the results of Rosenberg and McKenny (1983), who obtained 500 GalK units for this strong promoter. As the GalK units obtained with P_{27}, P_{31} and P_{CS3} were 99.1, 72.2 and 157.7, respectively, it indicates that these three ETEC promoters are of moderate strength (Table 5.1).

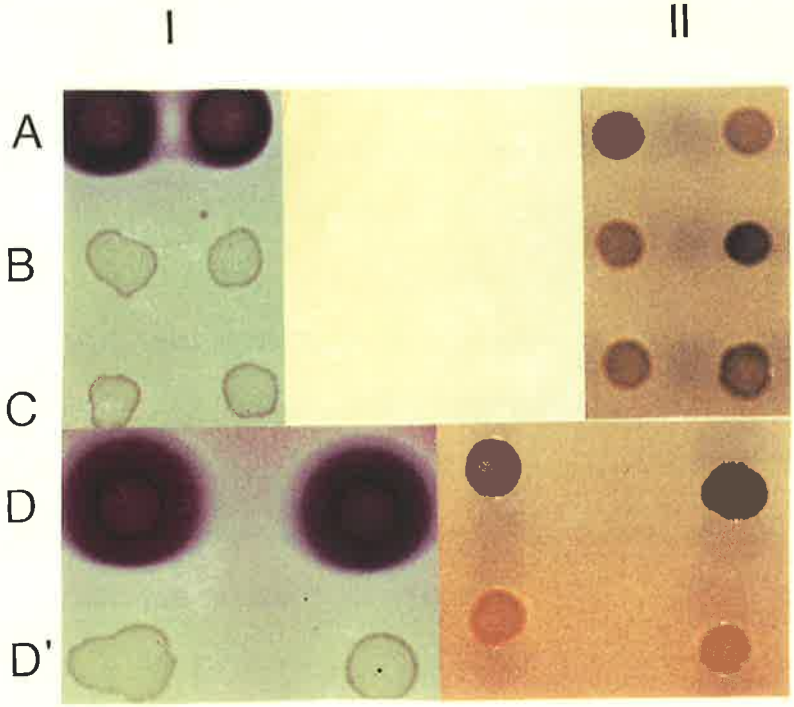
Table 5.1

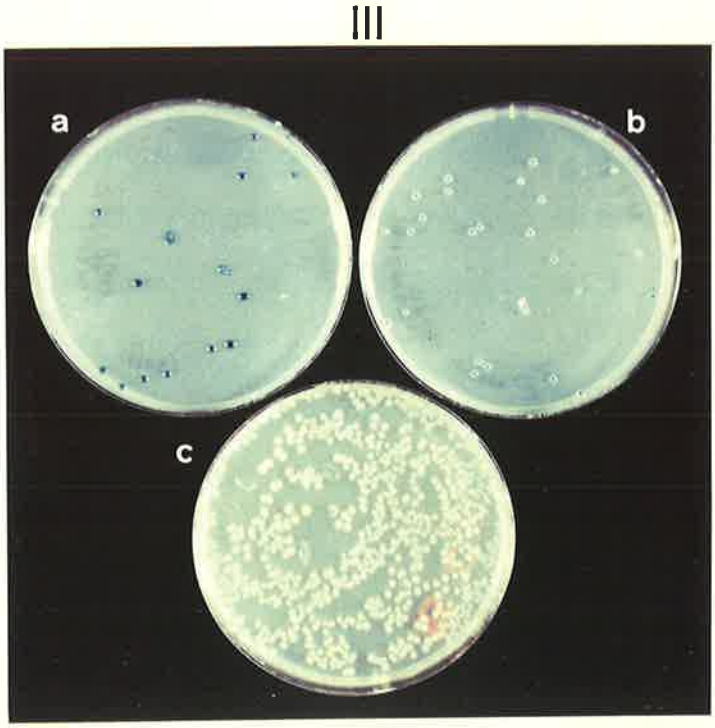
Galactokinase Assay to detect promoter activities

Fragment	Putative Promoter	GalK Units
A	P_{27}	99.10
B	P_{61}	0.35
C	P31	72.20
D	P_{CS3}	157.70
Vector Control	--	2.86
Positive Control	P_{lac}	620.00

Figure 5.6 Detection of promoter activities *in vivo*

Bacteria carrying the pCB clones containing P_{27}, P_{61}, P_{31} and P_{CS3} (see figure 5.2 for map) were tested on galactose or lactose indicator plates. The plates were photographed under ordinary light. Groups A-D represent colonies of cells carrying the pCB clones in which the fragments are cloned in the appropriate direction to promote the transcription of the *galK* or *lacZ* gene. Group D' carries the P_{CS3} (fragment D in Figure 5.2) in the direction opposite to the *galK* and *lacZ* genes. Cultures in D (II) and D' (II) are plated for single colonies, alongwith control culture carrying pCB182, in panel III: *a* - P_{CS3} in the lacZ direction, *b* - P_{CS3} in the galK direction (i.e. testing the reverse PCS3) and *c* - pCB182.


I
II
A
B
C
D
D'


III
a
b
c

P_{27} mRNA, but is translated at a lower efficiency, the low intensity of the 4.3 kb band could result from a degradation of the mRNA in the absence of translation.

The translation of *cp61* through *cp22* could be from the P_{61} and/or P_{31} transcripts. Since a single transcript, initiating from P_{27}, covered the entire CS3 gene cluster, it could function as an operon even if no other promoter was present on the DNA. These additional promoters, including the strong P_{CS3}, are probably required for the production of the correct stoichiometric amounts of each of the proteins. Taylor *et al.* (1984) have suggested that internal promoters play a major role in achieving discoordinate regulation of genes within operons during special or extreme conditions.

The overlapping transcription in the CS3 gene cluster further confirms the arrangement of the genes. As previously predicted (Manning *et al.*, 1985), the DNA included in pPM484 is too short to encode all genes in a conventional, non-overlapping fashion, and therefore the bacterium is probably maximising its use by a combination of transcriptional and translational overlapping. Further studies, aimed at *trans*-complementation of mutations in different genes, will help to understand the functional significance, if any, of this arrangement. The mutagenesis of downstream genes by frameshift, nonsense or missense mutations will affect the product of *cp82*, one of the essential genes for pilus biogenesis. Locating the exact promoters and altering their -10 regions, without affecting the amino acid sequence within *cp82*, should help to ascertain the specific roles of these internal promoters as well as of the protein products, Cp61 through Cp9.

CHAPTER 6

Translational Control Of CS3 Pilus Biosynthesis

CHAPTER 6

Translational Control Of CS3 Pilus Biosynthesis

6.1 Introduction

In order to produce the 22 kDa polypeptide seen in minicells, the translation initiating from the AUG at nt 3530 must continue past the amber codon at nt 3578. Critical examination of the sequence did not reveal any error that would suggest an alternative reading frame to account for this polypeptide (Figure 6.1). The codon usage plot indicated that this region of the DNA (Figure 3.5) was probably coding and, furthermore, the plot showed good coding potential across this UAG codon in the same reading frame (section 3.2.4.3). If a sequencing error had occurred in this region, which resulted in a frameshift near the UAG codon, it should have been indicated by a shift in the codon usage plot. The absence of any frameshift error in this region was also confirmed by showing that the Cp9 polypeptide increased in size as predicted, when the upstream UAG codon (UAG_{3523}) was mutated to a sense codon to allow continued translation up to UAG_{3578} (Figure 4.5, lane *c*). For these reasons, a non-conventional translation for *cp22*, by read-through of UAG_{3578}, was predicted. As no precedent for this mode of translation of pilus biosynthesis genes was found in the literature, it was decided to investigate this phenomenon in some detail. This chapter describes the preliminary work towards verifying this hypothesis by investigating the effect of amber suppression on the production of CS3 pili.

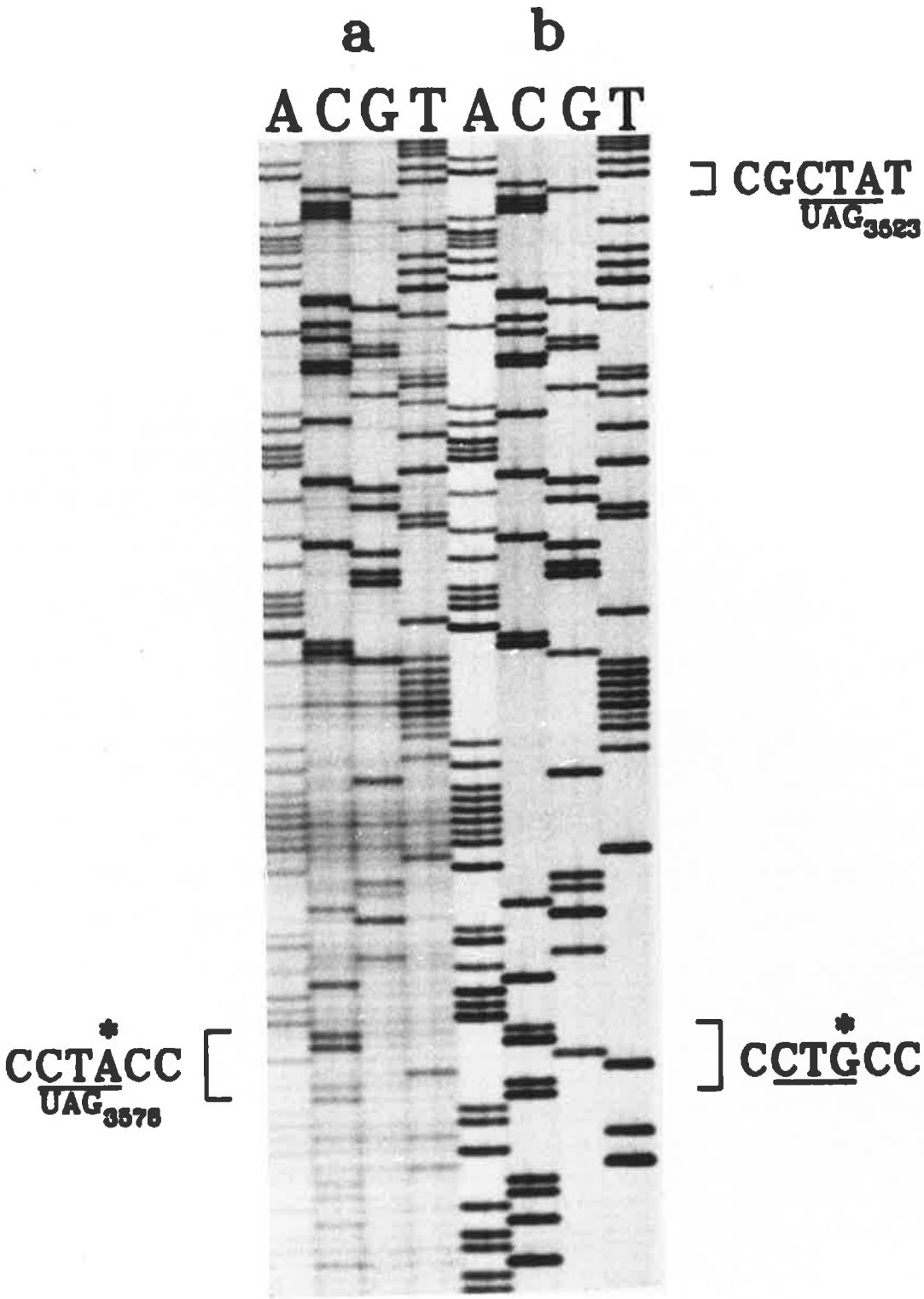
6.2 Results

6.2.1 Suppression of UAG_{3578} is required for CS3 biosynthesis.

The minicell-producing strain (DS410), which was routinely used in labelling the polypeptides and to verify pilus biosynthesis, carries a *supE* mutation (Manning, P.A. and Achtman, M., unpublished data). The plasmid, pPM484, directs the synthesis of pili in

Figure 6.1 Nucleotide sequence between UAG_{3523} and UAG_{3578}

The sequence, determined from an M13 clone carrying the oligomutagenized UAG_{3578}, is complementary to that shown in Figure 3.4a. Panel ***a*** gives the wild type sequence for comparison.Panel ***b*** shows the sequence between the two UAGs in the clone carrying the mutated UAG_{3578}. Asterisks indicate the nucleotide that was changed by oligomutagenesis.


a
b
A C G T A C G T
CGCTAT
UAG3523
CCTACC
UAG3575
CCTGCC

this strain (Figure 3.1 and 3.2). When this plasmid was introduced into two isogenic minicell-producing strains, PC2251 (Su^+) and PC2251 (Su^-), pilus biosynthesis was detected only in the Su^+ host. This was verified by immunogold EM analysis and Western blotting of crude pili preparations (Figure 6.2 and 6.3, lanes *i* and *j*). This implied that suppression of the UAG triplet in the reading frame of Cp22 was required for pili formation. Since the wild type *E. coli* K-12 is Su^- (i.e. *sup*$^+$), and if the same is true for ETEC, then pili cannot be formed in PB176. It appears, therefore, that either the wild type ETEC strains are Su^+ or a chance mutation, during or after the subcloning of DNA into pPM484, has resulted in the occurrence of a UAG triplet on the DNA. If the latter was true, then, re-introducing the DNA from pPM484 into a $CS3^-$ derivative of PB176, namely $PB176p^-$, should not restore pilus biosynthesis. To test this, the *Hin*dIII fragment from pPM484 was cloned into a mobilizable vector, pSUP202 (Simon *et al.*, 1983), and introduced into the ETEC strain $PB179p^-$ by conjugation. This derivative expressed CS3 pili, as verified by Western blotting and immunogold EM (Figure 6.4 and 6.5). This eliminated the possibility of a chance mutation in pPM484. To see whether the suppression of the amber codon in Su^+, but not in its isogenic Su^- derivative, was a peculiarity of the strain, PC2251, the plasmid was introduced into an Su^- *E. coli* K-12 strain, R594 (Bachman, 1987). Normal pili synthesis was observed in this strain, as evidenced in crude pili preparations (data not shown). This suggested that some mechanism other than amber suppression was involved in the read-through in this case. Whether the amber codon is suppressed in ETEC as in PC2251 (Su^+) or as in R594 is unknown at this stage.

6.2.2 Mutating UAG_{3578} to a sense codon restored pilus biosynthesis.

In a *supE* strain, the UAG amber codon is recognised by the tRNA for Gln (CAG codon; Parker, 1989; Fox, 1987). If this amino acid is being inserted at UAG_{3578} in PC2251 Su^+, then changing the UAG codon to a CAG triplet should result in pili synthesis in PC2251 (Su^-). This change was achieved by oligonucleotide mutagenesis

Figure 6.2 Immunogold EM of PC2251 (Su^+) and PC2251 (Su^-) minicells

The minicells were labelled with rabbit anti- CS3 pilin antibody followed by proteinA-colloidal gold complexes.

A - Su^+ minicell harbouring pPM484

B - Su^- minicell harbouring pPM484

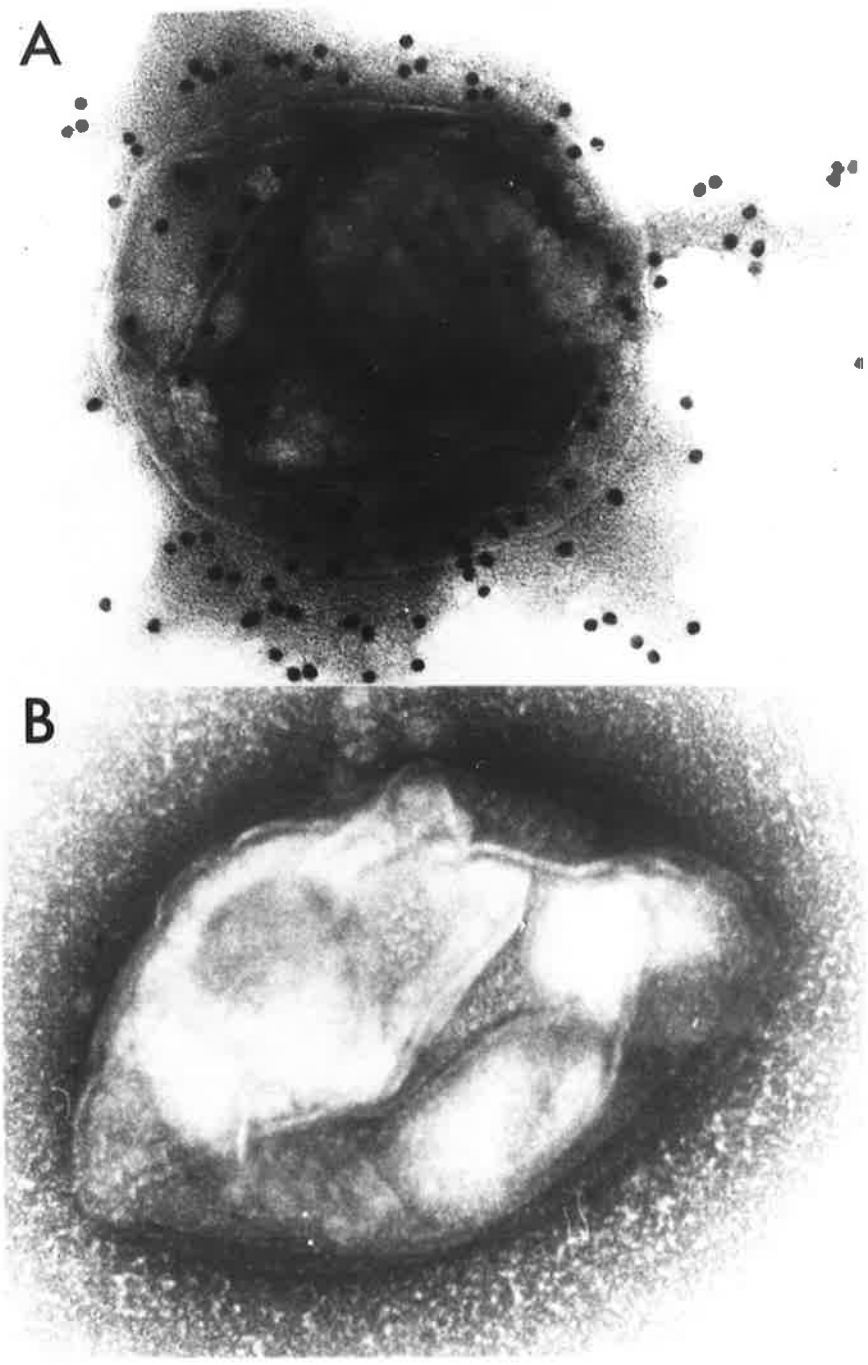

A
B

Figure 6.3 Western blot of crude pili from Su^+ and Su^- strains:

Crude pili were prepared from PC2251 (Su^+) and PC2251 (Su^-) carrying pPM484 or its mutant derivatives, run in SDS on 15% PAGE, transferred to nitrocellulose and probed with CS3 antiserum.

Lanes:-

a - Su^- (pBR322)

b - Su^+ (pBR322)

c - Su^- (UAG_{3578} mutant)

d - Su^+ (UAG_{3578} mutant)

e - Su^- (*Sac*I mutant)

f - Su^+ (*Sac*I mutant)

g - Su^- (UAG_{3523} mutant)

h - Su^+ (UAG_{3523} mutant)

i - Su^- (pPM484)

j - Su^+ (pPM484)

k - $PB176p^-$

l - PB176

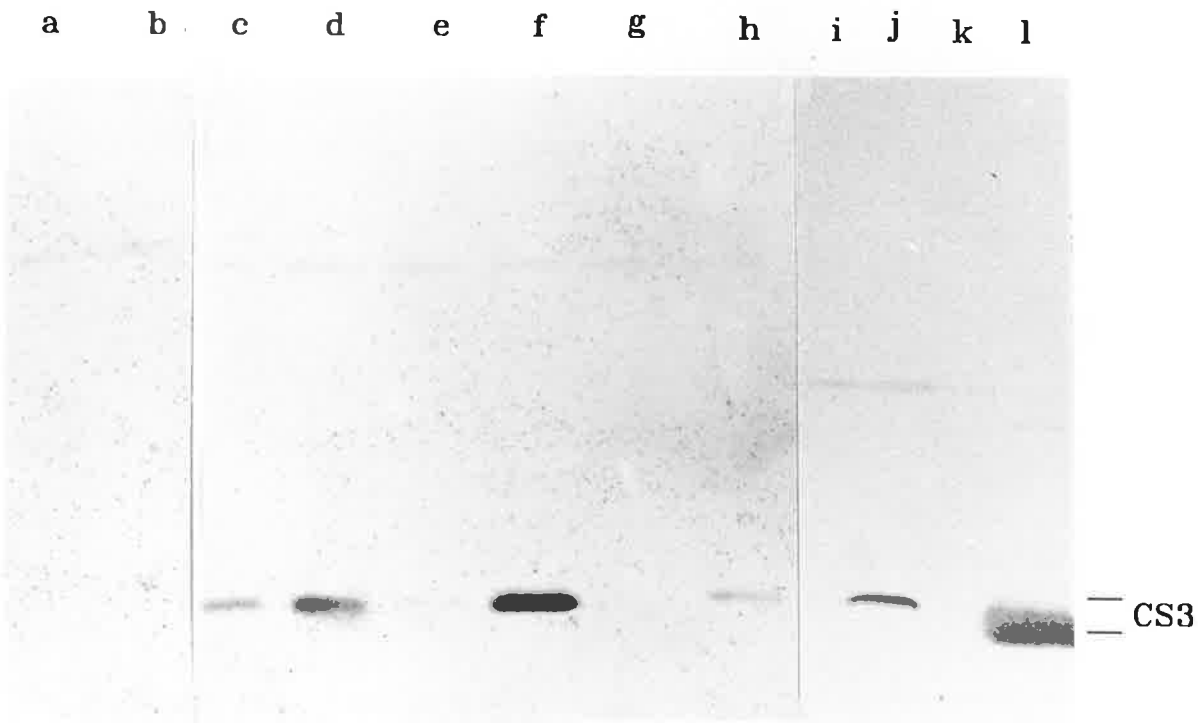

a
b
c
d
e
f
g
h
i
j
k
l
CS3

Figure 6.4 Western blot of crude pili from *E. coli* K12 and ETEC strains

Crude pili from the respective strains were run in SDS on 15% PAGE, transferred to nitrocellulose and probed with CS3 antiserum.

Lanes:

a - pSUP202 in SM10

b - pPM744 in SM10

c - pPM744 in PB176p$^-$

d - PB176p$^-$

e - PB176

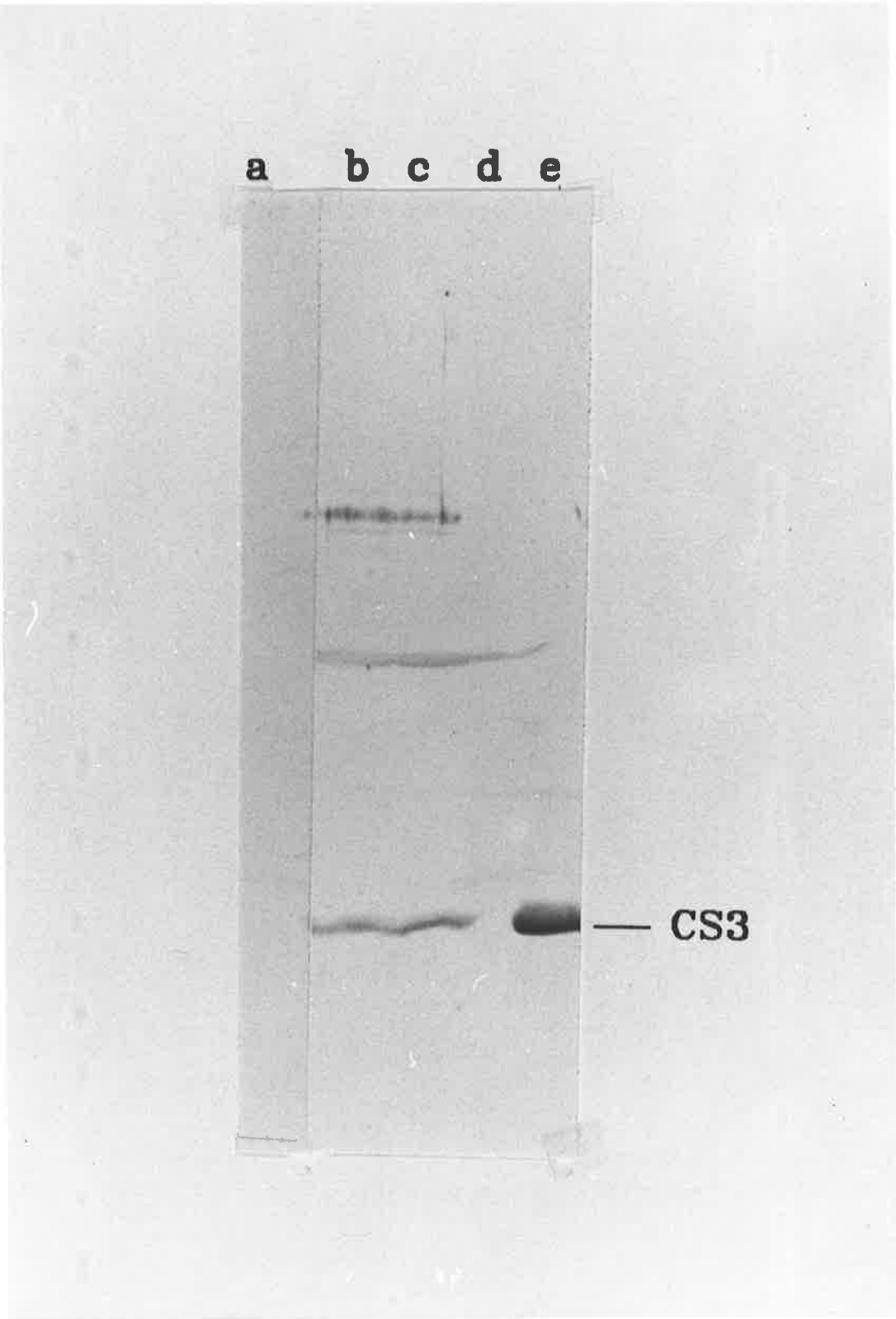

a b c d e
CS3

Figure 6.5 Immunogold EM of ETEC, mobilized with DNA from pPM484

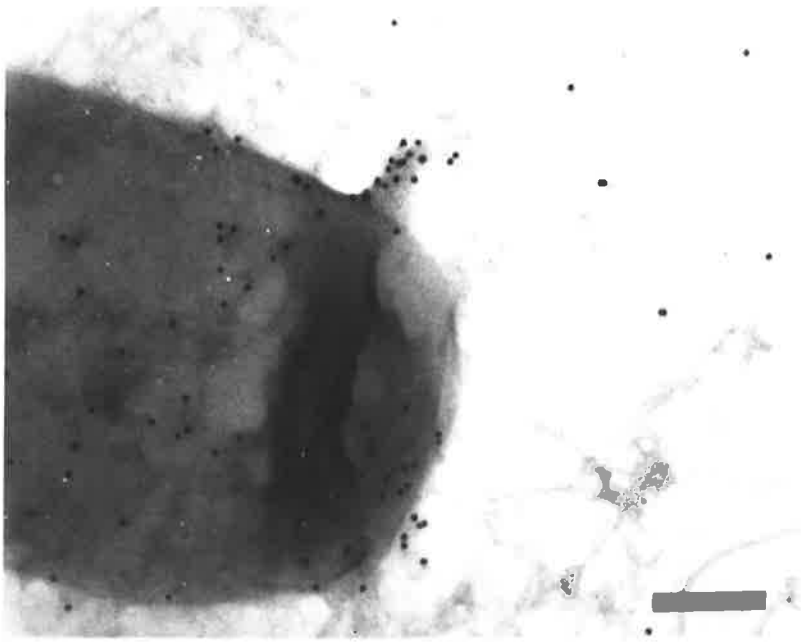
The cells were labelled with rabbit anti-CS3 pilin antibody followed by protein-A -colloidal gold complexes.

a - PB176p$^-$ harbouring the plasmid, pPM744

b - PB176p$^-$

Magnification bars indicate 0.37μM in both cases.

a



b

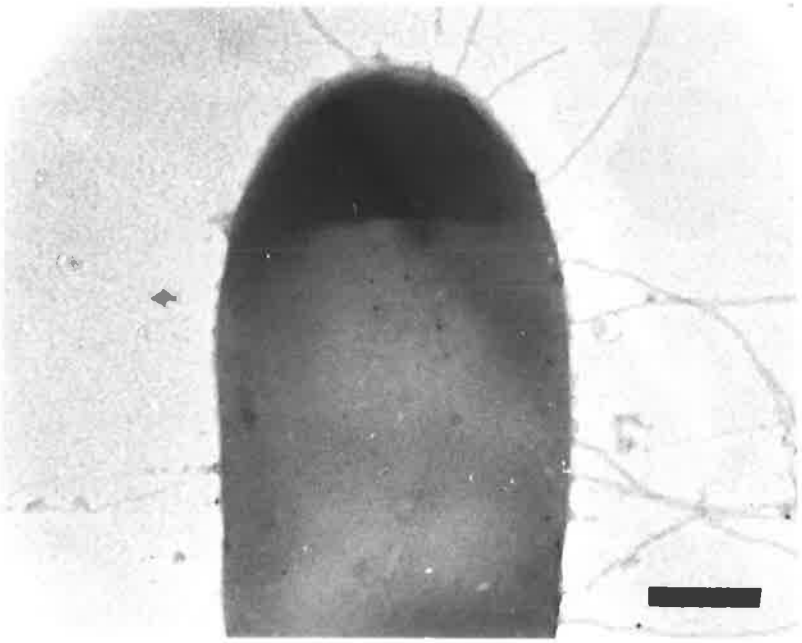


Figure 6.6 Immunogold EM of UAG mutants

The cells were labelled with rabbit anti-CS3 pilin antibody followed by protein-A -colloidal gold complexes. Lengths of the magnification indicators are shown in brackets.

a - UAG_{3578} mutant in (Su^+) (0.18μM)

b - UAG_{3578} mutant in PC2251 (Su^-) (0.24μM)

c - UAG_{3523} mutant in DS410 (0.24μM)

d - *Sac*I mutant in DS410 (0.30μM)

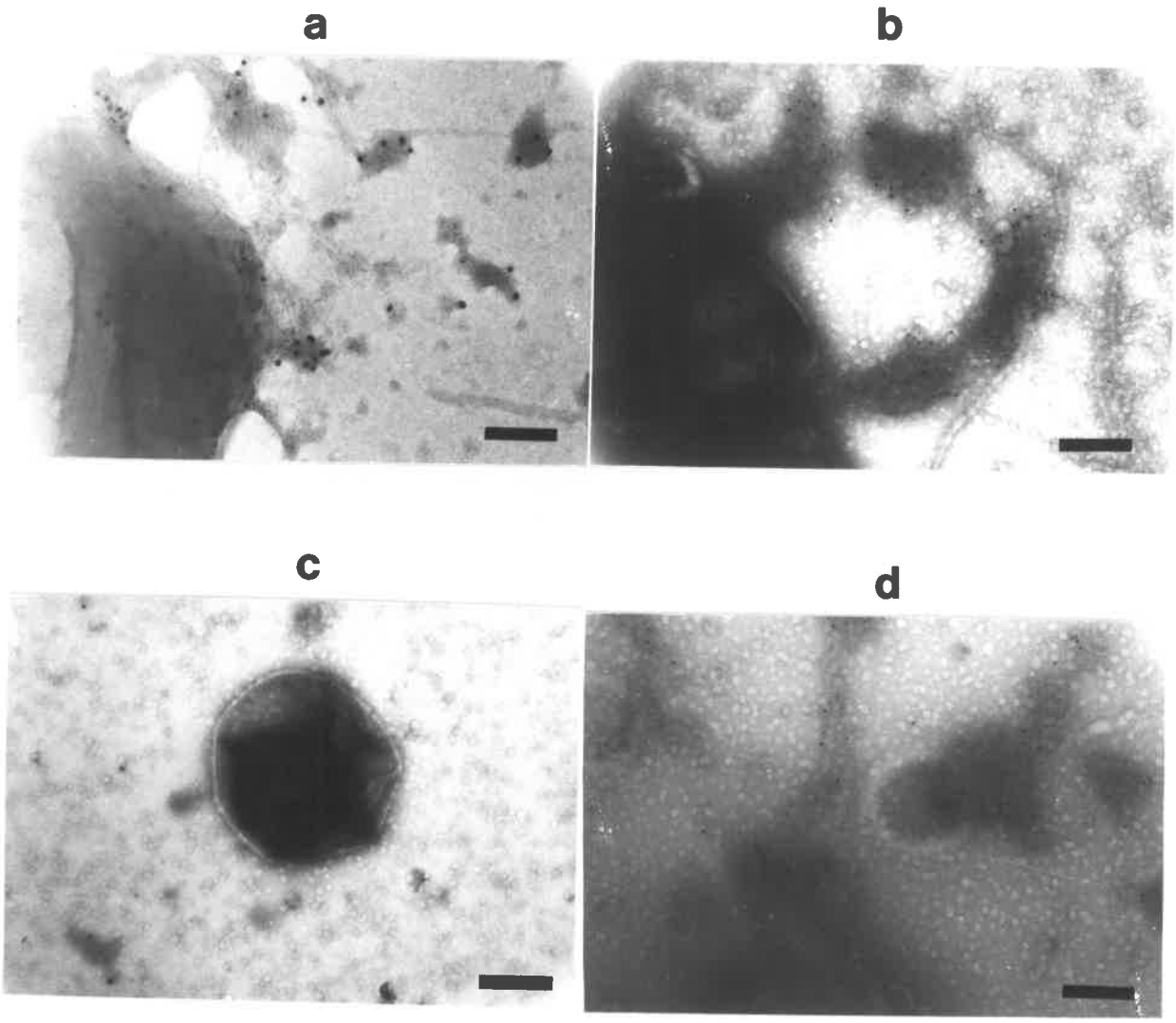

a
b
c
d

Figure 6.7 Western blot of crude pili from *supF* strain carrying pPM484 DNA

Crude pili from the respective strains were run in SDS on 15% PAGE, transferred to nitrocellulose and probed with CS3 antiserum.

a - pPM4026 in PC2251 (Su^+)

b - pPM4026 in PC2251 (Su^-)

c - CDM8 in PC2251 (Su^+)

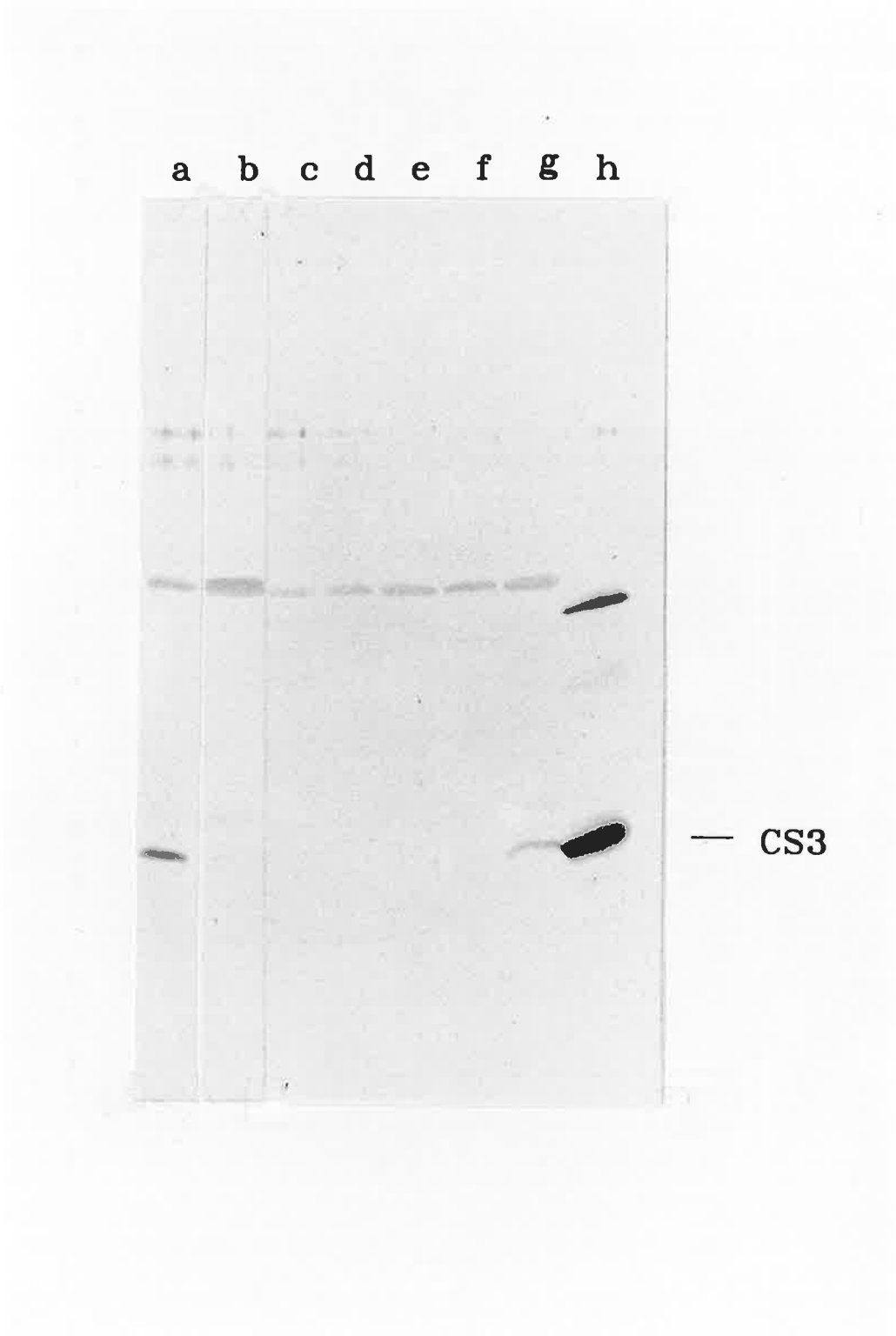
d - CDM8 in PC2251 (Su^-)

e - pPM4027 in PC2251 (Su^-)

f - pPM484 in PC2251 (Su^-)

g - pPM484 in PC2251 (Su^+)

h - PB176


a b c d e f g h
CS3

(Figure 6.1, panel *b*), and pilus synthesis was monitored in PC2251 (Su^+) and PC2251 (Su^-) strains. Western blotting of crude pili preparations from the transformants (Figure 6.3, lanes *c* and *d*), as well as immunogold EM (Figure 6.6 a and b), showed that pili were produced in both strains. As the wild type pPM484 and the mutant plasmid differed only in a single base pair, C-G, instead of T-A at position 3578, it indicated that insertion of Gln at the position of the amber codon could restore pilus biosynthesis.

6.2.3 Mutating UAG_{3523} to a sense codon resulted in loss of pilus biosynthesis.

In order to ascertain whether read-through of UAG_{3523} has any effect on pilus biosynthesis, it was mutagenised to a GAG triplet. This resulted in loss of pilus synthesis as indicated by immunogold EM (*see* Figure 6.6c). In this case, however, a very faint band corresponding to CS3 subunit was seen in Su^+, but not in Su^-, background by Western blotting (*see* Figure 6.3, lanes *g* and *h*). Interestingly, end-filling of the newly created *Sac*I site at UAG_{3523} returned the pilus synthesis capacity to the cells (*see* Figure 6.3, lanes *e* and *f* ; Figure 6.6d).

6.2.4 Suppression of UAG_{3578} by *supF* did not restore pilus biosynthesis.

The suppressor gene, *supF* (*tyrT*), inserts the amino acid Tyr at a UAG termination codon (Gallucci and Garen, 1966). The effect of this suppression on pilus synthesis was tested by introducing the cloned *supF* gene into a strain harbouring pPM484. This was accomplished in two steps: firstly, the 4.7 kb *Hin*dIII fragment from pPM484 was cloned into a plasmid pMac5-8 (P. Stanssens, unpublished data), which contained an amber mutation in the β-lactamase gene. This was then introduced into PC2251 (Su-) harbouring a Km^R plasmid carrying the cloned *supF* gene (pPM827, derived from CDM8; Seed, 1987). The transformants now became Ap^R, owing to the *supF*-mediated suppression of the amber mutation in the β-lactamase gene, but no CS3 pili were detected in immunoblots of crude pili preparations (Figure 6.7, lane *e*). Thus,

although suppression was occurring, it appeared that the Tyr residue inserted at UAG_{3578} by means of the *sup*F suppressor was unsuitable for restoring pilus synthesis.

6.3 Discussion

The results presented in this chapter imply that suppression of the UAG triplet in *cp22* is essential to produce cell-surface pili. When tested in isogenic Su^+ and Su^- strains, this suppression appeared to require the insertion of a particular amino acid (Gln), as indicated by the absence of pilus production when suppressed by the insertion of Tyr by *supF*. However, in wild type ETEC and in R594, where pili are formed in the absence of suppression, a read-through rather than insertion of an amino acid is presumably occurring. This region of Cp22 is probably critical for its maturation, but insertion of certain amino acids or the omission of a residue may be tolerated. A membrane associated segment has been predicted in Cp22 between the residues 8 and 22, and may form part of the signal sequence. Since the amino acid (residue 17) inserted at the amber codon would lie within this region, it is possible that specific residue at this position could affect accessibility by the signal peptidase and not permit mature Cp22 to be made. A computer analysis, with all possible amino acid residues in this position, showed that the substitution of Tyr for Gln, as well as omission of a residue, actually increased the probability of a trans-membrane segment. Without a clearer understanding of the role of Cp22, it is difficult to speculate how, and whether, the presence of a trans-membrane segment in it affects pili synthesis.

Although amber codons usually occur as a result of mutations in *E. coli* K-12, use of these codons as a regular mode of control of gene expression occurs in many RNA viruses (Parker, 1989). RNA dependent DNA polymerase (reverse transcriptase) is apparently a product of misreading at the amber codon in the *pol* gene of the RNA virus, murine leukemia virus (Shinnick, *et al.*, 1981). The termination at the leaky UAG stop codon results in the synthesis of a 126 kDa protein of tobacco mosaic virus (Pelham,1978), whereas read-through (at an efficiency of 10%) leads to the synthesis of

a 183 kDa product essential for infectivity (Ishikawa, *et al.*, 1986). Chang *et al.* (1989) suggest that read-through translation may produce polypeptides whose amino terminal peptides serve as leader sequences as well as independent polypeptides.

Suppression in an Su^+ host only produces about 10% of the normal level of polypeptide (Atkins *et al.*, 1990). This can conceivably work as a modulator of expression of some genes. Compared to the amounts of other proteins visualised by minicell translations, the product of *cp22* always appeared as a very faint band, supporting the proposition that the inframe UAG codon is a modulator of its expression. In an *in vitro* translation system, however, this protein was synthesised at an elevated level. This could reflect either an elevated level of tRNAs in the *in vitro* system, or the accumulation of Cp22 in the absence of pilus assembly.

Translations of *cp82, cp61, cp46, cp31* and *cp9* terminate at UAG_{3523} (Figure 4.5). No significant suppression seems to occur at this codon, as indicated by the intensity of bands. Although a faint band in lane *d* of Figure 4.5 is of the same length as the Cp9 product after the mutagenesis of UAG_{3523} (Figure 4.5, lane *c*), it is most probably an artefact, rather than read-through of Cp9. This probably signifies that suppression, or read-through, of an amber codon is highly context-dependent and that the sequence around the first UAG triplet does not permit it.

It is also significant that translation of UAG_{3523}-mutated Cp9, as well as of the other upstream proteins, stopped at, rather than read past UAG_{3578} (Figure 4.5, lane *c*). It appears to be an enigma that requires further study. Functional significance of an inframe UAG triplet in *cp22* may also be revealed by further studies. Cp22 shows significant homology to the PapH protein, and hence may serve as a basal protein or length modulator in CS3 pili synthesis. Perhaps an unusual control in its translation may serve to keep its expression at the required levels.

CHAPTER 7

General Discussion

CHAPTER 7

General Discussion

7.1 Introduction

Pilus biosynthesis requires the co-ordinate expression of a number of genes. In addition to the pilin subunit proteins, the pilus operons consist of genes encoding regulatory, transport, assembly and structural proteins. Five to eleven polypeptides have been implicated in the phenotypic expression of *E. coli* fimbriae (Mooi *et al.*, 1982; Norgren *et al.*, 1984; Van Die *et al.*, 1984a; De Graaf *et al.*, 1984; Båga *et al.*, 1985; Hultgren *et al.*, 1991). Co-ordinate expression of proteins is accomplished in the simplest way by having the genes which code for such proteins organised within a transcriptional unit, an operon, under the control of one promoter that is regulated by signals in the environment. The classical example is the *E. coli lac* operon (Jacob and Monod, 1961). The pilus biosynthesis genes are clustered together and in most cases present as a single transcriptional unit (Mooi *et al.*, 1981; Dougan *et al.*, 1983; De Graaf *et al.*, 1984; De Graaf, 1990). Though putative promoters were observed upstream to the first three genes (*fanA*, *fanB* and *fanC*) in the K99 operon, the remaining five genes were not preceded by a promoter. It appears therefore that the whole operon is controlled by the regulatory region upstream of *fanC* (De Graaf, 1990). The same appears to be true in the case of K88, TypeI and Pap systems (De Graaf, 1990), with a difference in that the transcription of *papI* proceeds in the direction opposite to that of the rest of the genes. Although co-ordinate expression of genes is generally achieved at the level of transcription (Rosenberg and Court, 1979), differential expression of genes within an operon, resulting in different stoichiometry of the gene products, is obtained *via* regulation that acts at levels beyond the initiation of transcription.

Studying the organisation of genes involved in CS3 pilus biogenesis was important in addressing the question of whether the general features of pili operons were retained in this system. Previous work (Manning *et al.*, 1985) has shown that a minimum

coding region of at least 3.75 kb of DNA was needed for the expression of CS3. Since this was far larger than the length required to encode a 15 to 17 kDa major subunit (approx. 450 nt), it re-affirmed that more than the pilus subunit was required for CS3 pilus biogenesis. At least seven polypeptides were observed in minicell translations using plasmids carrying this minimal DNA, and deletion analyses suggested that all were needed to express cell-surface pili. The observation, that the length of the minimal DNA was well short of the length estimated from the sizes of these polypeptides, led to the proposal that some of these polypeptides could be either derived by post-translational processing of larger molecules, or encoded as overlapping translational units (Manning *et al.*, 1985). If the latter were true, then regulation of expression at the level of translation could be occurring in CS3 pilus biogenesis.

The aim of this thesis was to provide a better understanding of the organisation of the genes involved in the biosynthesis of CS3 fimbriae and to assign possible roles to the various protein products.

7.2 Gene organization within the CS3 operon

Transcription, as indicated by RNA Northern blots and by various promoter-detection methods, appears to originate from at least four sites within the minimal DNA (Figure 3.6). The highest activity was found for the promoter in front of CS3, followed by that in front of *cp27* and *cp31* (Table 5.1). Another promoter in front of *cp61* has also been detected by Northern blot analysis, and colorimetrically on BCIG plate. Apparently, expression of the genes involved in CS3 pilus biogenesis is not from a single transcript, and in this respect differs from that in other pilus systems. The nomenclature 'CS3 gene cluster', instead of 'CS3 operon', is therefore used to signify the possibility that different gene products may be synthesised from different mRNAs. Sequence analysis revealed that eight potential ORFs were identified on the cloned DNA and five of these could be correlated with the protein products visualised in earlier minicell studies (Manning *et al.*, 1985). Three additional proteins, observed during the

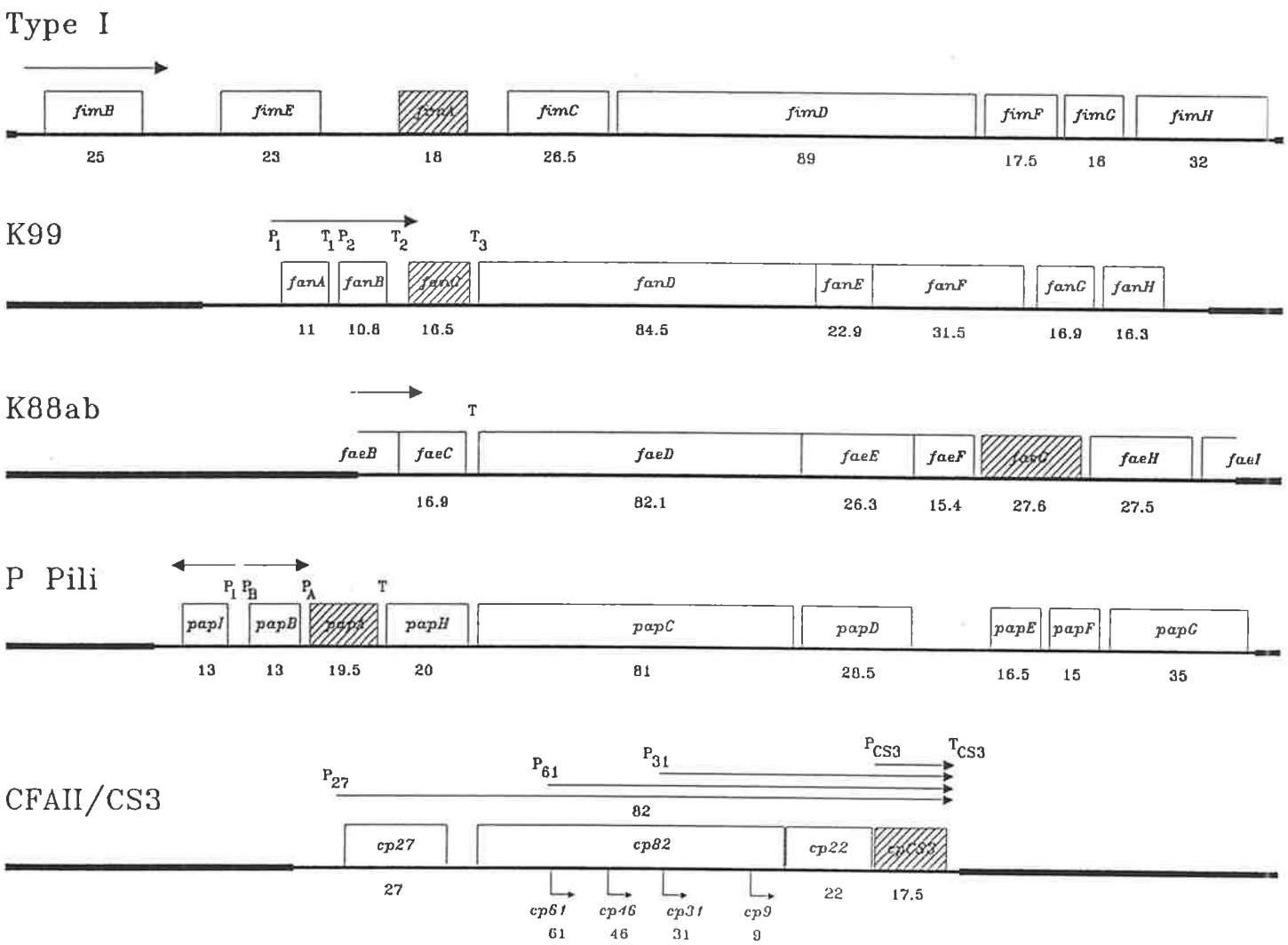
course of the present study, were confirmed to be present on the plasmid (pPM474) used by Manning *et al.* (1985). Though the number of genes within the CS3 gene cluster was similar to that in other pilus systems, their organisation was different. Firstly, the spatial organisation of genes on the DNA was quite different from that in Type I, Pap, K88 and K99 systems (De Graaf, 1990; Figure 7.1). The relative location of the genes for the putative channel protein (*cp82*) and periplasmic transporter (*cp27*) in the CS3 gene cluster is opposite to that found in the Pap, K88 and K99 systems (De Graaf, 1990). Only the Type I pilus offers some similarity in the gene arrangement, as the equivalent genes, *fim*C and *fim*D (Klemm and Christiansen, 1987; De Graaf, 1990), are arranged in the same order as *cp27* and *cp82*. In all the above pilus systems, the genes are arranged contiguously but do not overlap, except for minor translational coupling. In the case of CS3 pilus system, four genes (*cp27*, *cp82*, *cp22* and *cpCS3*) are similarly placed - contiguous, but not overlapping with each other (Figure 3.6). Another four ORFs (*cp61*, *cp46*, *cp31* and *cp9*), however, were completely contained within *cp82*. Finally, a very intriguing feature is the presence of an inframe UAG triplet within one of the genes, *cp22*. Suppression, or some other form of bypassing of this codon, appears to be an integral part of CS3 regulation. These differences, especially the possible overlapping of translation and the inframe UAG codon, make the CS3 gene cluster unique among all other pilus systems so far studied.

7.3 Possible significance of overlapping genes

Since the most striking feature of the CS3 gene cluster was the overlapping organisation of the genes, the main emphasis of the present study was to identify the translation initiation sites of these polypeptides and to assign possible roles to their proteins. Even though post-translational processing and degradation of polypeptides used are conceivable alternatives to produce the polypeptides visualised in minicell translations, site-directed mutagenesis unambiguously proved independent initiations of translation. The direct involvement of these proteins in pilus synthesis could not be

Figure 7.1 Structural organization of genes involved in fimbrial biogenesis

This figure has been adapted from De Graaf (1990), by adding the map of the CS3 gene cluster for comparison. Genes are represented by boxes, with the predicted sizes (kDa) of their polypeptides marked beneath. Positions of the overlapping genes in the CS3 gene cluster are indicated by short arrows. Thick lines represent the vector sequences. Transcripts are indicated by thin lines with arrow heads. Known or predicted promoters and terminators are marked. Genes for the pilin subunits are hatched.


Type I
fimB
fimE
fimA
fimC
fimD
fimF
fimG
fimH
25
23
18
26.5
89
17.5
16
32
K99
P1
T1 P2
T2
T3
fanA
fanB
fanC
fanD
fanE
fanF
fanG
fanH
11
10.8
16.5
84.5
22.9
31.5
16.9
16.3
K88ab
T
faeB
faeC
faeD
faeE
faeF
faeG
faeH
faeI
16.9
82.1
26.3
15.4
27.6
27.5
P Pili
P1 PB
PA
T
papI
papB
papA
papH
papC
papD
papE
papF
papG
13
13
19.5
20
81
28.5
16.5
15
35
CFAII/CS3
P27
P61
P31
PCS3
TCS3
82
cp27
cp82
cp22
cpCS3
27
22
17.5
cp61
cp46
cp31
cp9
61
46
31
9

demonstrated due to the concomitant effect of all mutations, within their coding regions, on Cp82. This raises the question as to whether they are actually needed for pilus synthesis. It is possible that these proteins are artefacts seen in the minicell expression system, and may not be produced in wild type ETEC at all. Visualisation of these bands, by ^{35}S-Met labelling, in the ETEC strain, PB176, is impractical and immunoprecipitation with antibodies to these proteins may be the only way to identify them. However, the following observations suggest that these proteins are real and are probably needed for pilus biosynthesis in the wild type ETEC. Firstly, the transcription pattern in PB176 is similar to that in *E. coli* K-12, with four bands of the same length being detected in Northern blots (section 5.2.2, Figure 5.3). The finding that these transcripts all have the same polarity, as well as length in both organisms, strongly suggests that they are identical. The estimated locations of two promoters, upstream of *cp61* and *cp31*, suggested that these two genes, and possibly *cp46* and *cp9* as well, are encoded on transcripts that are different to the one encoding *cp82*. Thus, even though these genes share the same reading frame with *cp82*, they are transcribed and translated independently. Consequently, their proteins are almost certainly present in PB176 as well as in *E. coli* K-12.

Secondly, the truncation of Cp61 and Cp46 by the *Bgl*II deletion resulted in an accumulation of Cp22 (Figure 4.3A, lane *j*). The *Nco*I mutation that affected Cp82 without affecting the other two polypeptides did not result in this accumulation (Figure 4.3A, lane *f*). This suggests that the products of *cp61* and *cp46* play a role in the utilisation of Cp22 in pilus synthesis. The above observations support that these proteins are indeed part of the pilus biosynthesis apparatus.

Although overlapping translation has a distinct advantage to viruses and bacteriophages (Parker, 1989; Normark *et al.*, 1983b), where the size of the genome is small, its use in CS3 fimbrial biosynthesis genes is surprising. It has been proposed that bacterial polycistronic mRNA is sequentially translated by the same ribosomes from the

5'-proximal cistrons to those distal to it (Ames and Hartman 1963; Danchin and Ullmann, 1980; Brot *et al.*, 1981). Ribosomes can also re-initiate internally in a cistron after encountering a nonsense mutation (Files *et al.*, 1974; Parker, 1989). It is believed that in these situations the ribosome release factor somehow fails to cause the dissociation of the translating ribosomes from the mRNA, so that they now scan the downstream region for an initiating Met residue (Parker, 1989). It seems possible that internal *de novo* initiations within a reading frame can also occur (Manley, 1978). The latter appears to be happening in the CS3 gene cluster, as no stop codon is encountered before any of the four internal initiation sites. Instead, five proteins with identical C-terminal ends are being formed by independent translation (section 4.2.2). One could, however, speculate that occasional frameshift during the translation of upstream genes could result in termination in the vicinity of the initiation codons of the downstream genes. The ribosomes may then re-initiate from the downstream AUG triplets. Such frameshifts are not uncommon and can be induced by amino acid starvation (Parker, 1989, Weiss and Gallant, 1986). A more likely scenario is that the genes in CS3 gene cluster are translated from different mRNAs. The appearance of the 2.9 and 2.0 kb mRNAs in the Northern blot favours this hypothesis. These RNAs start downstream to the initiation codon of *cp82*, and are probably used for the translation of *cp61* and *cp31*, respectively. Other, as yet undetected mRNAs may encode *cp46* and *cp9*.

Apart from independent translations, post-translational cleavage of the 82 kDa polypeptide could also, conceivably, produce the five different polypeptides observed in minicells. However, since there are several proteins to be made from a single precursor molecule, it is essential to ensure that a single molecule of the polypeptide is cut only once. Unless specific cleavage mechanisms operate to generate molecules with different N-termini, this method is likely to result in a complete degradation of the precursor to the smallest protein in the set. This mode of producing the smaller polypeptides will also be uneconomical, as the N-terminal peptides remaining after cleavage will have to be

discarded. In separate translations, however, only a required level of the polypeptides may be produced, and this can be both economical and efficient to the organism.

Perhaps one purpose of overlapping genes in the CS3 gene cluster is to effect a stoichiometric expression of the various proteins needed for pilus synthesis. By having independent transcription and/or translation to produce these polypeptides, the bacterium can probably ensure the synthesis of appropriate, but different concentrations of the various proteins.

7.4 Pattern of Transcription

It appears that more than one promoter is involved in the transcription of the genes for CS3 pilus biosynthesis. All transcripts terminate after the CS3 coding region, as indicated by the ability of CS3 probes to pick up all mRNA bands (Figures 3.6 and 5.3). The largest transcript originates upstream from *cp27* and probably encodes all eight genes. Transcripts initiating from promoters within *cp82*, but upstream from *cp61*, *cp31* and *CS3*, may encode the latter and other genes contained within them. Since multiple initiation sites can be used to generate different proteins from the same mRNA, as explained in the previous section, multiple transcripts to encode the downstream genes may also signify a different level of control for their expression. The existence of internal promoters in *FSend*A (endoglucanase gene from *Fibrobacter succinogenes* AR1) suggested that gene expression could be modulated by the individual promoters. Cavicchioli and Watson (1991) suggested that repression of the primary transcript and derepression or induction of the secondary transcript could control expression of an operon, of which *FSend*A is a part. Taylor *et al.* (1984) have suggested that internal promoters play a major role in achieving discoordinate regulation of genes within operons during special or extreme conditions. Thus, a single mRNA might encode all genes under normal conditions of growth, but under a different set of conditions (e.g. change in growth temperature or nutritional state) the downstream promoters might become active. The expression of pili is affected by growth conditions such as

temperature and composition of the growth medium (Gaastra and De Graaf, 1982; De Graaf *et al.*, 1980). In the case of Pap pili, transcription of the structural gene for the pilus subunit (*papA*) is regulated in response to the growth temperature (Göransson and Uhlin, 1984). Transcription of *papB*, and *papA* has been shown to be under catabolite repression (Båga *et al.*, 1985). Temperature also appears to be an important factor to which this transcriptional regulatory system is responsive, as *papI* has been implicated in the thermoregulation of Pap fimbrial expression (Göransson *et al.*, 1989). Investigation of the pattern of transcription in the CS3 gene cluster under different growth conditions might reveal any possible significance of these internal promoters in regulating CS3 expression.

7.5 Codon Usage

In unicellular organisms, codon choice bias is often correlated with the level of the tRNA species that reads the codon and is most pronounced in highly expressed genes (Ikemura, 1985). Kurland (1987) has postulated that use of a limited number of codons, and hence a limited number of tRNAs, for highly expressed genes is a strategy for optimising translational efficiency. Marked differences in the codon usage preferences seem to exist between *E. coli* K-12 and the ETEC pilus biosynthesis genes. This was originally indicated by the observation that the codon usage table based on *E. coli* K-12 genes did not predict the coding regions in pPM484 DNA, whereas one based upon other ETEC pilus genes readily detected the potential coding areas. Closer examination of the codon usage tables (Table 7.1) revealed that some of the codons predominantly used in ETEC were used at a low frequency in *E. coli* K-12. Specifically, the triplets AUA (for Ile), ACA (for Thr) CGG and AGA (for Arg) and GGA and GGG (for Gly) are rare codons in *E. coli* K-12 but these are used at the same frequency as others in ETEC. In five out of nine cases, where only two codons existed for the same amino acid, the preferences were opposite in *E. coli* K-12 and ETEC. The codon usage in *cp82* showed similarity to that in ETEC genes, concomitant with an even more evident

Table 7.1

Codon frequencies in ETEC, *E. coli* K-12 and the *cp* genes

aa	Codon	ETEC	*E.coli*	Cp82	Cp27	Cp22	CS3
F	UUU	67.81	43.69	90.91	77.78	72.73	100.00
F	UUC	32.19	56.31	9.09	22.22	27.27	0.00
L	UUA	16.40	8.50	41.51	24.00	31.25	40.91
L	UUG	9.46	10.06	13.21	4.00	12.50	18.18
S	UCU	27.23	32.78	28.85	62.50	25.00	35.71
S	UCC	22.28	32.50	7.69	12.50	16.67	7.14
S	UCA	30.69	14.22	50.00	12.50	41.67	50.00
S	UCG	19.80	20.50	13.46	12.50	16.67	7.14
Y	UAU	64.88	47.67	64.81	42.86	80.00	0.00
Y	UAC	35.12	52.33	35.19	57.14	20.00	100.00
*	UAA	70.00	71.5	0.00	0.00	100.00	100.00
*	UAG	10.00	7.0	100.00	0.00	100.00	0.00
C	UGU	75.76	42.14	100.00	66.67	0.00	0.00
C	UGC	24.24	57.86	0.00	33.33	0.00	0.00
*	UGA	20.00	21.5	0.00	100.00	0.00	0.00
W	UGG	100.00	100.00	100.0	100.00	100.00	0.00
L	CUU	12.30	8.20	28.30	20.00	25.00	18.18
L	CUC	5.05	9.54	1.89	8.00	6.25	0.00
L	CUA	3.47	2.39	11.32	36.00	6.25	9.09
L	CUG	53.31	61.32	3.77	8.00	18.75	13.64
P	CCU	21.19	12.50	14.29	33.33	66.67	40.00
P	CCC	16.56	7.35	9.52	6.67	0.00	20.00
P	CCA	17.88	17.63	71.43	40.00	33.33	40.00
P	CCG	44.37	62.52	4.76	20.00	0.00	0.00
H	CAU	67.44	44.31	80.00	50.00	14.29	50.00
H	CAC	32.56	55.69	20.00	50.00	28.57	0.00
Q	CAA	18.06	28.00	47.83	50.00	42.86	0.00
Q	CAG	81.94	72.00	52.17	50.00	14.29	50.00
R	CGU	47.33	50.28	7.41	11.11	0.00	100
R	CGC	19.85	37.94	7.41	0.00	40.00	0.00
R	CGA	9.92	3.86	7.41	11.11	0.00	0.00
R	CGG	22.90	5.11	3.70	0.00	0.00	0.00
I	AUU	49.05	41.29	51.85	70.83	84.62	72.73
I	AUC	20.00	55.10	18.52	0.00	0.00	0.00
I	AUA	30.95	3.61	29.63	29.17	15.38	27.27
M	AUG	100.00	100.00	100.00	100.00	100.00	100.00
T	ACU	17.24	21.19	35.00	21.43	37.50	37.04
T	ACC	27.59	48.56	17.50	21.43	12.50	18.52
T	ACA	31.03	9.24	45.00	42.86	37.50	37.04
T	ACG	24.14	21.00	2.50	14.29	12.50	7.41
N	AAU	60.98	32.26	77.05	72.22	72.73	83.33
N	AAC	39.02	67.74	22.95	27.78	27.27	16.67
K	AAA	66.49	76.05	82.22	72.73	92.31	75.00
K	AAG	33.51	23.95	17.78	27.27	7.69	25.00
S	AGU	64.10	26.28	88.46	85.71	72.73	66.67
S	AGC	35.90	73.72	11.54	14.29	27.27	33.33
R	AGA	72.50	1.75	51.85	55.56	20.00	0.00
R	AGG	27.50	1.06	22.22	22.22	40.00	0.00
V	GUU	30.07	31.79	50.00	38.46	64.71	66.67
V	GUC	10.51	17.59	8.33	0.00	5.88	0.00
V	GUA	23.91	17.73	27.78	46.15	17.65	11.11
V	GUG	35.51	32.89	13.89	15.38	11.76	22.22
A	GCU	21.37	20.24	33.33	27.27	30.00	37.50
A	GCC	26.34	23.29	20.00	0.00	30.00	12.50
A	GCA	31.30	21.55	30.00	63.64	30.00	37.50
A	GCG	20.99	34.92	16.67	9.09	10.00	12.50
D	GAU	60.28	54.20	74.42	72.73	83.33	100.00
D	GAC	39.72	45.80	25.58	27.27	16.67	0.00
E	GAA	59.49	71.40	62.50	76.92	66.67	75.00
E	GAG	40.51	28.60	37.50	23.08	33.33	25.00
G	GGU	33.60	42.77	28.38	41.18	53.33	55.56
G	GGC	19.47	41.06	16.22	5.88	13.33	11.11
G	GGA	22.93	6.18	33.78	29.41	26.67	11.11
G	GGG	24.00	9.99	21.62	23.53	6.67	22.22
Total		3920	71735	736	242	200	169

disparity with the *E. coli* K-12 genes. In addition to the rare codons mentioned above, the triplets UUA (for Leu) and AGG (for Arg) were found at relatively high levels in the *cp82* ORF. As the relative abundance of tRNAs can dictate the rate of translation (Ikemura, 1985), the presence of rare codons in the coding region can act to modulate the expression of genes. The *E. coli* K-12 codon usage table (Maruyama *et al.*, 1986) was based on a random selection of genes whose sequences were known, without any distinction as to the type or level of expression of these genes. The inability of this table to predict the coding region of the CS3 genes may stem from the fact that the genes for pilus biosynthesis, in general, belong to a class of proteins expressed at a lower level and therefore use a specific set of codons under-represented in the table. Expression of these genes in *E. coli* K-12 led to pili on the cell surface, which suggested that the presence of rare codons did not affect the production of the required amounts of proteins in this bacterium. Assuming that no fundamental difference in codon usage exists between the two organisms, the use of rare codons in the pili synthesis genes could be a form of regulation of their expression in ETEC.

7.6 Read-through Translation

In Chapter 3, an assumption was made that the reading frame of *cp22* passed across a UAG termination codon. Site-directed mutagenesis proved that *cp22* was indeed located in this region. The oligonucleotide mutagenesis that removed UAG_{3523}, where the upstream genes terminated, resulted in lengthening of the Cp82, Cp61, Cp46, Cp31 and Cp9 polypeptides in agreement with the prediction that they now terminated at UAG_{3578}. This showed that the latter UAG triplet, in the reading frame of *cp22*, did not occur as a result of a sequencing error. A chance mutation in the plasmid, pPM484, that resulted in an inframe UAG triplet was ruled out when the same area of the DNA was sequenced from the independently isolated plasmid, pPM474. Furthermore, re-introduction of the minimal DNA for pilus synthesis into a deletion derivative of PB176 resulted in restoration of pilus synthesis. It was therefore assumed that

suppression or read-through translation of the UAG codon was essential for production of Cp22 and, thereby, pilus synthesis. Analyses in isogenic Su^+ and Su^- minicell strains, where pili were formed only in the Su^+ strain, supported this assumption. In wild type ETEC, expected to be Su^-, some other form of read-through must be occurring. Genetic background of the strain must be playing role in read-through, as implied by the observation that one Su^- *E. coli* K-12 strain (R594) produced pili when transformed with the plasmid pPM484. Programmed read-through of nonsense codons have been observed in many instances (Parker, 1989). One of the best-characterised example is in tobacco mosaic virus (TMV). The most 5' open reading frame encodes a 126 kDa protein and terminates at a UAG (Goelet *et al.*, 1982). Reading this termination codon would allow the synthesis of a 183 kDa protein, which is seen in both *in vitro* and *in vivo* (Paterson and Knight, 1975; Pelham, 1978). In addition to TMV, the turnip yellow mosaic virus (Haenni, *et al.*, 1987) and the beet yellow vein virus (Bouzoubaa *et al.*, 1986) show read-through translation at a UAG. A wide variety of other plant RNA viruses also seem to produce read-through proteins (Haenni, *et al.*, 1987; Harbison *et al.*, 1985; Hsu and Brakke, 1985; Hughes, *et al.*, 1986; Ziegler, *et al.*, 1985). The use of UAG and UGA codons is also common in the attenuators of *E. coli* (Parker, 1989). In the above examples the read-through occurs naturally *in vivo*, but the amount of such read-through could be increased *in vitro* by suppressor tRNAs. It has been postulated that the use of read-through could simply ensure that a protein is made in low abundance or targeted to a particular location (Parker, 1989). The N-terminus of the protein in these instances could serve as a leader sequence as well as an independent polypeptide (Chang *et al.*, 1989). In the case of Cp22, however, such an independent peptide that terminated at UAG_{3578} would be very short (16 aa), and therefore its detection in minicell translations is not practical.

One paradox must be noted in accepting the suppression/read-through hypothesis to explain the synthesis of Cp22: if suppression occurs at the downstream UAG triplet,

then, it must also occur at the UAG codon that marks the end of translation of the upstream genes. If this were the case, extended products of *cp82* through *cp9* must be visible in minicell translations, but these were not present. Suppression of amber codons is influenced by the sequence context in which they are found (Ryoji *et al.*, 1983; Atkins *et al.*, 1990). It is possible that the sequence around the downstream UAG triplet somehow favours suppression or bypassing of this codon, whereas that around the upstream UAG triplet prevents such read-through.

The control of expression of *cp22* is probably unique in more than one way. The sequence near its AUG initiation codon does not show an RBS and the ribosomes terminating at the UAG_{3523} triplet, which occurs 7 bases upstream to the initiation codon of *cp22*, may be essential for creating a localised concentration and thereby facilitating re-initiation. Perhaps this serves as a mechanism to translationally couple the expression of *cp22* with the upstream genes.

7.7 Properties of proteins

Examination of the secondary structure profiles of the Cp proteins indicates that all are devoid of extensive α-helices but rich in β-sheets. Marked hydrophobic regions occur along the length of Cp82.

The predicted isoelectric points of the proteins varied widely, with Cp27 having the highest (pI = 9.9) and Cp9 having the lowest (pI = 4.53). The pI values of other ETEC proteins that showed homology to Cp27 were also high (e.g. PapD = 9.77, FanE = 9.17, FaeE = 10.23). Similarly, the pI values of Cp82 and other channel proteins showed some similarity (e.g. Cp82 = 6.42; PapC= 6.44; FanD = 6.14; FaeD = 5.60).

CS3 lacks cysteine residues, which suggests that non-covalent forces are responsible for the structural integrity of the native protein. Other fimbrial proteins lacking cysteine residues include *cfaA* (Klemm, 1982; Hamers *et al.*, 1989), CS2 (Klemm, *et al.*, 1985a), CS1 (Hall *et al.*, 1989), K88ab (Gaastra *et al.*, 1981) and antigen 8786 (Aubel *et al.*, 1991). Since disulphide bonds may be important for protein-protein

interactions, and are of critical importance in a structural protein, the morphogenesis of CS3 fimbriae may be significantly different from that of Pap and Type I fimbriae. The C-terminal portion of the CS3 polypeptide, like that of other fimbrial subunits, is hydrophobic. It has been proposed that such segments may be involved in intersubunit interactions which maintain the integrity of fimbrial superstructure (De Graaf and Mooi, 1986; Klemm, 1985).

Although the mature CS3 product lacked any Met residue, if processing occurred after the Ala (residue 22) as indicated by the N-terminal sequence, this polypeptide was being labelled with ^{35}S-Met in minicell translations. An alternative processing site after the residue 14, in which case one Met residue will be retained in the mature protein, has been demonstrated. This presumably forms the band migrating slower than the major pilus subunit from PB176 and has been observed in minicell translations. Processing apparently occurs at both these sites in PB176, as indicated by the presence of two bands in crude pili preparations from this strain. Only one band, corresponding to the larger, less abundant component is seen in the minicells. This suggests that only one processing site, the one after residue 14, is being used in *E. coli* K-12. Since pili were being formed in *E. coli* K-12, it seems that an extra 7aa at the N-terminal end does not affect its function. Whether this alternative processing is a normal function in the wild type ETEC, or an artefact created by difference in the signal peptidase specificity in *E. coli* K-12 is not known.

7.8 Cellular location of gene products

Cell fractionation clearly showed that the mature Cp27 was entirely located in the periplasm. Although none of the mature form was seen in the membrane fractions, the precursor form was found in the membrane, as might be expected for molecules in the process of being translocated across the inner membrane. Homology to other chaperone proteins in the ETEC pilus systems makes Cp27 a clear candidate for being a periplasmic transporter/chaperone protein. Contrary to the expectation, the mature form of CS3 was

not found in the membrane fractions, but some was found in the periplasm. This probably represents the pool of molecules waiting to be transported to the outer membrane. It seems likely that the shearing forces exerted during fractionation could have dislocated the fimbriae on the cell surface so that no CS3 was seen in the outer membrane. As for the other six proteins in the CS3 system, all appeared to be membrane associated. Although the fractionation of minicell membranes is fraught with problems (Achtman *et al.*, 1979) it seems clear that most of the proteins are found in the inner membrane. This is very different to what has been observed in other ETEC and Pap pili systems. These proteins most likely have structural roles, and therefore might be predicted to be located in the outer membrane, however, with the possible exception of Cp22 they do not possess discernible leader peptides. The same applies to Cp82. It seems more likely that these proteins span the inner membrane. Although the experiments in this thesis do not conclusively resolve the localizations of the proteins, it is possible that the use of gene fusions based on *lacZ*, *phoA* or *bla* may provide genetic evidence for their membrane topology and imply inner or outer membrane localization (Manoil and Beckwith, 1985; Broome-Smith and Spratt, 1986).

If the interpretation of the data is correct, then, how are CS3 fimbriae assembled using a series of proteins which appear to be cytoplasmic membrane-located? A possible clue comes from studies of secretion of the haemolysin (HlyA) of uropathogenic *Escherichia coli* (Blight and Holland, 1990). In this case, the HlyA molecule, which lacks a signal peptide but possess a specific secretion signal within the last 50 C-terminal amino acids, is transported directly to the medium by a translocation complex formed by HlyB and HlyD (Gray *et al.*, 1986; Mackman *et al.*, 1986; Koronakis *et al.*, 1989). Both HlyB and HlyD have been located to the inner membrane of *E. coli* (Mackman *et al.*, 1985; Holland *et al.*, 1989) but they fractionate in small but significant amounts with the outer membrane fractions, similar to the proteins seen here. It has been proposed that they form a complex which spans both membranes, allowing direct secretion of HlyA

into the medium (Blight and Holland, 1990). A similar situation could be visualized in the CS3 system, where the products of Cp61, Cp46, Cp31 and Cp9 forming a complex which translocates Cp82, the channel protein, to the outer membrane. Another system of proteins that do not possess signal sequences, but which are outer membrane located, are the Ipa invasion/adhesion proteins of *Shigella* (Venkatesan *et al.*, 1988). These proteins are assembled on the cell surface in a signal-sequence-independent manner. Perhaps such a system also applies to CS3.

The interpretation of the role of Cp31 is somewhat difficult. It does not have a discernible signal sequence and appears to migrate aberrantly on SDS-PAGE - sometimes faster than *pre-bla* and at other times slower. Such anomalies have previously been observed in *E. coli* and are, in part, related to the concentrations and purity of SDS (Manning and Reeves, 1977). A 30 kDa protein is thought to function in association with the periplasmic transporter protein to prevent preliminary polymerisation of the pilin subunits and to keep the major and minor subunits in a conformation favourable for translocation across the outer membrane (De Graaf, 1990). It also possesses affinity to the pore protein and is the first protein to be inserted in the fimbrial growth point. Although Cp31 does not show significant homology to this 30 kDa FimH protein (Klemm and Christiansen, 1987), it has homology to the putative pore protein and similarity in size to FimH. The presence of a separate promoter for this gene (*cp31*) may also be significant.

Cp22 appears to have a signal peptide, and its homology to PapH (Båga *et al.*, 1987) strongly suggests that it is a minor pilin component. Cp9 was also clearly found in the membrane fraction, and is, therefore, likely to play a structural role in CS3 biogenesis.

7.9 Model for CS3 pilus biosynthesis

Based on the predicted roles of Cp27, Cp82, Cp22 and CS3, and the intracellular location of their products, the following model for CS3 pilus biogenesis, diagramatically

given in Figure 7.2, may be proposed: Independent transcription and translation of the genes, *cp27* through *cpCS3*, produces the various polypeptides. The control of expression of the polypeptides is probably regulated at the level of transcription and translation, but none of the genes on the minimal DNA appear to have a role in the expression of other genes. The physiological condition of the cells may play an important role, and this is somewhat suggested by the fact that pili are not formed at 30°C (De Graaf *et al.*, 1980; Mooi *et al.*, 1986; Roosendaal, 1987) and that differences in the level of piliation exist between cells grown in different media.

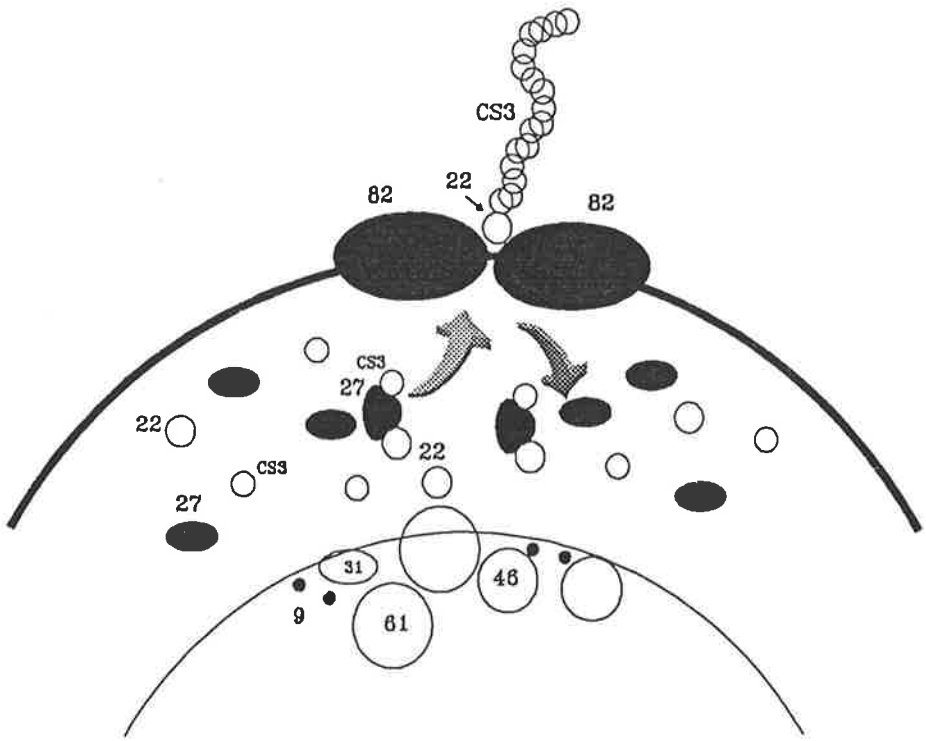
Once in the periplasm, the CS3 polypeptide associates with Cp27, which modifies its structure and/or transports it to the outer membrane. In the Pap pilus system, the PapD protein, to which Cp27 shows homology, is proposed to act at the level of transport of pilins to the outer membrane (Båga *et al.*, 1987), and is also needed for stabilisation of non assembled pilins in the periplasm (De Graaf, 1990). In the Type I pilus system, the 30 kDa FimH protein is associated with the periplasmic transporter, FimC, at this stage (Klemm and Christiansen, 1987). Probably, in the CS3 system, Cp31 may be associated with Cp27 and the pili subunit. Additionally, this complex is likely to contain Cp22, which probably acts as the basal protein during the assembly of pili. The similarity of Cp22 to PapH (Båga *et al.*, 1987) suggests that, like the latter, this protein may form the basal part of the pili itself and helps in the cell-anchorage of pilus structure.

Cp82 may form channels in the outer membrane through which CS3 passes to form the cell-surface pili. The poorly expressed PapC was found to be located in the outer membrane (Norgren *et al.*, 1987) where it forms channels in the outer membrane through which pilins are transported and assembled into the pilus structure. A similar function has been proposed for FaeD in the K88 system (Mooi *et al.*, 1983), FanD in K99 (Roosendaal and De Graaf, 1989) and FimD in type I pili systems (Klemm and Christiansen, 1990).

The remaining three proteins (Cp61, Cp46 and Cp9) may play structural roles, in

Figure 7.2 Model for CS3 pilus biosynthesis

Diagramatic representation of the predicted locations and roles of the gene products. Shown are the inner (thin line) and outer (thicker line) membranes and the periplasmic space. Proteins are represented by ovals and circles, roughly in proportion to the relative sizes of polypeptides. Possible protein-protein associations are indicated by grouping the symbols together. Even though the active form of proteins may contain more than one subunit, this is not assumed in the figure. Arrows indicate the Cp27-mediated transport of pilin subunits to the outer membrane.


CS3
82
22
82
CS3
27
22
22
CS3
27
31
46
9
61

conjunction with Cp82, in forming the channel. Alternatively, they may be involved in the transport of Cp82 to the outer membrane. The accumulation of Cp22 in a *cp61/cp46* mutant suggests a possible interaction between either or both these proteins and the product of *cp22*. These proteins may also play a role in the Cp27-mediated transport of CS3 to the membrane, as an accumulation of the CS3 precursor form was also seen in the above mutants. A possible scenario is that unless the mature CS3 is delivered to the outer membrane and assembled into pili, the Cp22 product will remain un-utilised.

In order to test the above model the protein complexes that may be formed during CS3 pilus biogenesis must be isolated intact and analyzed on PAGE under denaturing and non-denaturing conditions. Antibodies against CS3 might be useful in the precipitation of such protein complexes.

7.10 Conclusion

This thesis has sought to provide a better understanding of the molecular organisation and functioning of the CS3 pilus system in ETEC. In this regard, the minimal fragment required for pilus synthesis in *E. coli* K-12 was analysed, both by computer analysis of the DNA sequence and *in vivo* expression/manipulation of the proteins it encodes. The sequence and its analyses led to some surprises, the most important of which was the overlapping translation and an amber suppression as possible modes of control in pilus synthesis. Both of these are unique features of CS3 compared to all other pili systems studied so far, although ample evidence for such non-conventional control mechanisms exists in the expression of genes in viruses and bacteriophages. Since CS3 pili do not possess morphological characters totally different to other types of pili, this novel method of regulation of the genes involved in its synthesis is intriguing. From a standpoint of economy of the coding DNA, the organisation and expression of the genes in the CS3 gene cluster rivals even that found in some bacteriophages and RNA viruses, where such economy is vital. The CS3 genes, like some other pili gene systems, are located on a plasmid in the wild type DNA. Size

limitations of plasmids cannot be assumed to contribute a selection pressure on the operon sizes. One may speculate that the CS3 gene cluster was once part of a transposon or a bacteriophage which was integrated into one of the plasmids. The presence of part of the IS*91* element in the minimal DNA strongly points to this possibility.

Further studies on the regulation, especially that involving the amber suppression may be rewarding. In this respect, a region encoding the start site of *cp22* and the two UAG triplets could be fused to a reporter gene and expressed *in vivo* and *in vitro*. Site-directed mutagenesis of this construct may help to unravel the sequences responsible for effecting suppression at one amber codon but not the other. It may also reveal the requirements, if any, of upstream translation on the efficiency of initiation of *cp22*. Selectively mutating a gene and then supplying its protein *in trans* must be performed to see the effect of individual proteins on pilus synthesis. One way of mutating an overlapping gene in the CS3 gene cluster, without affecting Cp82, is to alter its RBS in such a way that the amino acid sequence is unaltered. Identifying the RBS and specifically changing the sequences to render them non-functional must be tried first on small pieces of DNA fused to a reporter gene, and then introduced into the minimal DNA for pilus synthesis. The inactivation of the internal promoters should also be done in the same manner, in order to verify whether they play any essential role in pilus synthesis. Finally, monoclonal antibodies against peptides from the different proteins must be raised, and used in immunogold EM studies to identify the intracellular location of these proteins.

Fundamental knowledge gained from the studies on the expression of CS3 pili could be utilised in a practical application such as generating oral vaccines. In addition to expressing the CS3 gene cluster, to elicit an immune response to the CS3 pilus itself, it may be possible to replace part of the pilus subunit with other antigenic determinants or insert suitable epitopes in exposed domains. Identifying the minimum length of CS3 subunit required to form pili, and defining the surface exposed domains will be a prelude

to inserting other antigenic proteins into its coding region. It is hoped that the data presented in this thesis will be of some assistance in designing future experiments in this direction.

REFERENCES

Abe, C., Schmitz, S., Moser, I., Boulnois, G., High, N. J., Orskov, I., Orskov, F., Jann, B. and Jann, K. (1987). Monoclonal antibodies with fimbrial F1C, F12, F13 and F14 specificities with fimbriae from *E. coli* O4:K12:H$^-$. *Microb. Pathog.* **2**: 71-77.

Abraham, J. M., Freitag, C. S., Clements, J. R. and Eisenstein, B. I. (1985). An invertible element of DNA controls phase variation of type I fimbriae of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **82**: 5724-5727.

Abraham, S. N., Goguen, J. D., Sun, D., Klemm, P. and Beachy, E. H. (1987). Identification of two ancillary subunits of *E. coli* type I fimbriae by using antibodies against synthetic oligopeptides of *fim* gene products. *J. Bacteriol.* **169**: 5530-5535.

Abraham, S. N., Goguen, J. D. and Beachy, E. H. (1988a). Hyper adhesive mutant of type I fimbriated *Escherichia coli* associated with the formation of FimH organelles (fimbriosomes). *Infect. Immun.* **56**: 1023-1029

Abraham, S. N., Sun, D., Dale, J. B. and Beachy, E. H. (1988b). Conservation of the D-mannose adhesion protein among type I fimbriated members of the family *Enterobactericeae. Nature (London)* **336**: 682-684.

Achtman, M., Schwuchous, S. and Helmuth, R., Morelli, G. and Manning, P. A. (1978). Cell-cell interactions in conjugating *Escherichia coli*: Con$^-$ mutants and stabilisation of mating aggregates. *Molec. Gen. Genet.* **164**: 171-183.

Achtman, M., Manning, P. A., Edelbuth, C. and Herrlich, P. (1979). Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor *tra* cistrons in *Escherichia coli* minicells. *Proc. Natl. Acad. Sci. USA* **76**: 4837-4841.

Aiba, H., Adhya, S. and Crombrugghe, B. (1981). Evidence for two functional gal promoters in intact *Escherichia coli* cells. *J. Biol Chem.* **256**: 11905-11910.

Ames, B. N and Hartman, P. E. (1963). The Histidine Operon. *Cold. Spring. Harbor. Symp. Quant. Biol.* **28**: 349-356.

Anderson, S. (1981). Shotgun DNA sequencing using cloned DNAseI- generated fragments. *Nucleic. Acids. Res.* **9**: 3015

Atkins, J. F., Weiss, R. B. and Gesteland, R. F. (1990). Ribosome Gymnastics-Degree of Difficulty 9.5, style 10.0. *Cell* **62**: 413-423.

Aubel, D., Darfeuille, M. A. and Joly, B. (1991). New adhesive factor (Antigen 8786) on a human enterotoxigenic *Escherichia coli* O117:H4 strain isolated in Africa. *Infect. Immun.* **59**: 1290-1299.

Bachman, B. (1987). Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. p.1190-1219. *In* Neidhart, F. C., Ingraham, J. L., Low, B., Magasanik, B., Schaechter, M. and Umbarger, E. (ed.) "*Escherichia coli* and *Salmonella typhimurium,* cellular and molecular biology." American Society of Microbiology, Washington, D.C.

Båga, M., Normark, S., Hardy, J., O' Hanley, P., Lark, D., Olsson, O., Schoolnik, G. and Falkow, S. (1984). Nucleotide sequence of the PapA gene encoding the pap pilus subunit of human uropathogenic *Escherichia coli. J. Bacteriol.* **157**: 330-333.

Båga, M., Göransson, M., Normark, S. and Uhlin, B.-E. (1985). Transcriptional activation of a Pap pili virulence operon from uropathogenic *Escherichia coli. EMBO. J.* **4**: 3887-3893.

Båga, M., Norgren, M. and Normark, S. (1987). Biogenesis of *E. coli* Pap pili: PapH, a minor pilin subunit involved in cell anchoring and length modulation. *Cell* **49**: 241-251.

Båga, M., Göransson, M. and Normark, S. (1988). Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. *Cell* **52**: 197-206

Bakker, D., Vader, C. E. M., Roosendaal, B., Mooi, F. R., Oudega, B. and De Graaf, F. K. (1991). Structure and function of periplasmic chaperone-like proteins involved in the biosynthesis of K88 and K99 fimbriae in enterotoxigenic *Escherichia coli*. *Mol. Microbiol.* **5**: 875-886.

Bakker, D., Willemsen, P. T. J., Simons, L. H., Van Zijderveld, F. G. and De Graaf, F. K. (1992). Characterisation of the antigenic and adhesive properties of FaeG, the major subunit of K88 fimbriae. *Mol. Microbiol.* **6**: 247-255.

Barnes, W. M., Bevan, M. and Son, P. H. (1983). Kilo-sequencing: Creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. *Meth. Enzymol.* **101**: 98-122.

Beachey, E. H. (1981). Bacterial adherence: adhesin-receptor interactions mediating the attachment of bacteria to mucosal surfaces. *J. Infect. Dis.* **143**: 325-345.

Benton, W. D. and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. *Science* **196**: 180

Betley, M. J., Miller, V. L. and Mekalanos, J. J. (1986). Genetics of bacterial enterotoxins. *Annu. Rev. Microbiol.* **40**: 577-605.

Birnboim, H. C. and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic. Acids. Res.* **7**: 1513-1523.

Black, R. E., Merson, M. H., Huq, I., Alim, A. R. M. A. and Yunus, M. (1981). Incidence and severity of rotavirus and *Escherichia coli* diarrhoea in rural Bangladesh: Implication for vaccine development. *Lancet.* **i**: 141-143.

Blight, M. A. and Holland, I. B. (1990). Structure and function of haemolysin B, P-glycoprotein and other members of a novel family of membrane translocators. *Mol. Microbiol.* **4**: 873-880.

Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H. and Low, D. A. (1989). Phase variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. *EMBO. J.* **8**: 613-620.

Blyn, L. B., Braaten, B. A. and Low, D. A. (1990). Regulation of pap pilin phase variation by a mechanism involving differential Dam-methylation states. *EMBO. J.* **9**: 4045-4054.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., Boyer, H. W., Crosa, J. H. and Falkow, S. (1977). Construction and characterisation of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**: 95-113.

Bortolussi, R., Ferrieri, P. and Quie, P. G. (1983). Influence of growth temperature of *Escherichia coli* on K1 capsular antigen production and resistance to opsonisation. *Infect. Immun.* **39**: 1136-1141.

Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K. and Jonard, G. (1986). Nucleotide sequence of beet necrotic yellow vein virus RNA-2. *J. Gen. Virol.* **67**: 1689-1700.

Boylan, M. and Smyth, C. J. (1985). Mobilisation of CS fimbriae-associated plasmids of enterotoxigenic *Escherichia coli* of serotype O6: K15: H16 or H^- into various wild-type hosts. *FEMS Microbiol. Lett.* **29**: 83-89.

Boylan, M., Coleman, D. C. and Smyth, C. J. (1987). Molecular cloning and characterisation of the genetic determinant encoding CS3 fimbriae of enterotoxigenic *Escherichia coli. Microb. Pathog.* **2**: 195-209.

Boylan, M., Coleman, D. C., Scott, J. R. and Smyth, C. J. (1988a). Molecular cloning of the plasmid-located determinants for CS1 and CS2 Fimbriae of enterotoxigenic *Escherichia coli* of serotype O6:K15:H16 of human origin. *J. Gen. Microbiol.* **134**: 2189-2199.

Boylan, M., Smyth, C. J. and Scott, J. R. (1988b). Nucleotide sequence of the gene encoding the major subunit of CS3 fimbriae of enterotoxigenic *Escherichia coli. Infect. Immun.* **56**: 3297-3300.

Brinton, C. C. Jr. (1959). Non flagellar appendages of bacteria. *Nature (London)* **183**: 782-786.

Brinton, C. C. Jr. (1965). The structure, function, synthesis and genetic control of bacterial pili and a molecular model for DNA and RNA transport in Gram-negative bacteria. *Trans. N. Y. Acad. Sci.* **27**: 1003-1054.

Brinton, C. C. Jr., Gemski, P. Jr., Falkow, S. and Baron, L. S. Location of the piliation factor on the chromosome of *Escherichia coli*. (1961). *Biochem. Biophys. Res. comm.*, **5**: 293-298.

Broes, A., Fairbrother, J. M., Jacques, M. and Lariviere, S. (1988). Isolation and characterisation of a new fimbrial antigen (CS 1541) from a porcine enterotoxigenic *Escherichia coli* O8:KX105 strain. *FEMS Microbiol. Lett.* **55**: 341-348.

Broome-Smith, J. K. and Spratt, B. G. (1986). A vector for the construction of translational fusions to TEM β–lactamase and the analysis of protrein export signals and membrane protein topology. *Gene* **49**:341-349.

Brot, N., Caldell, P. and Weissbach, H. (1981). Regulation of synthesis of *Escherichia coli* ribosomal proteins L1 and L11. *Arch. Biochem. Biophys.*, **206**: 51-53.

Brown, M. C. M., Weston, A., Saunders, J. R. and Humphreys, G. O. (1979). Transformation of *E. coli* C600 by plasmid DNA at different phases of growth. *FEMS Microbiol. Lett.*, **5**: 219-222.

Burgess, M. N., Bywater, R. J., Cowley, C. M., Mullan, N. A. and Newsome, P. M. (1978). Biological evaluation of a methanol-soluble, heat-stable *Escherichia coli* enterotoxin in infant mice, pigs, rabbits and calves. *Infect. Immun.* **21**: 526-531.

Calderwood, S. B. and Mekalanos, J. J. (1987). Iron regulation of shiga-like toxin expression in *Escherichia coli* is mediated by the fur locus. *J. Bacteriol.* **169**: 4759-4764.

Calogero, R. A., Pon, C. L., Canonaco, M. A. and Gualerzi, C. O. (1988). Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. USA* **85**: 6427-6431.

Calva, E., Torres, J., Vazquez, M., Angeles, V., de la Vega, H. and Ruiz-Palacios, G. M. (1989). *Campylobacter jejuni* chromosomal sequences that hybridise to *Vibrio cholerae* and *Escherichia coli* LT enterotoxin genes. *Gene* **75**: 243-251.

Caron, J. and Scott, J. R. (1990). A rns-like regulatory gene for Colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. *Infect. Immun.* **58**: 874-878.

Caron, J., Coffield, L. M. and Scott, J. R. (1989). A plasmid encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **86**: 963-967.

Cavicchioli, R. and Watson, K. (1991). The involvement of transcriptional read-through from internal promoters in the expression of a novel endoglucanase gene *FSend*A, from fibrobactor succinogenes AR1. *Nucleic. Acids. Res.* **19**: 1661-1669.

Chang, L. -J., Pryciak, P., Ganam, D. and Varmus, H. E. (1989). Biosynthesis of the reverse transcriptase of hepatitis B viruses involves *de novo* translational initiation not ribosomal frame-shifting. *Nature, (London)* **337**: 364-368.

Chomczynski, P. and Sacchi, N. (1987). Single step method of RNA isolation by Acid Guanidinium Thiocyanate Phenol-Chloroform extraction. *Ana. Biochem.* **162**: 156-159

Chou, P. Y. and Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**: 45-148.

Clark, C. A., Heuzenroeder, M. W. and Manning, P. A. (1992). Colonization Factor Antigen CFA/IV (PCF8775) of human Enterotoxigenic *Escherichia coli*: Nucleotide sequence of the CS5 determinant. *Infect. Immun.* **60**: 1254-1257.

Clegg, S. and Pierce, J. K. (1983). Organisation of genes responsible for the production of mannose resistant fimbriae of a uropathogenic *Escherichia coli* isolate. *Infect. Immun.* **42**: 900-906.

Clegg, S., Hull, S., Hull, R. and Pruckler, J. (1985a). Construction and comparison of recombinant plasmids encoding type I fimbriae of members of the family *Enterobacteriacae. Infect. Immun.* **48**: 275-279.

Clegg, S., Pruckler, J. and Purcell, B. K. (1985b). Complementation analyses of recombinant plasmids encoding type I fimbriae of members of the family *Enterobacteriaceae. Infect. Immun.* **50**: 338-340.

Clegg, S. and Gerlach, G. F. (1987). Enterobacterial fimbriae. *J. Bacteriol.* **169**: 934-938.

Connell, T. D and Holmes, R. K. (1992). Molecular genetic analysis of ganglioside GD1b-binding activity of *Escherichia coli* Type IIa heat labile enterotoxin by use of random and site directed mutagenesis. *Infect. Immun.* **60**: 63-70.

Cornelis, G., Sluiters, C., Lambertde Rouvroit, C. and Michiels, T. (1989). Homology between VirF, the transcriptional activator of the *Yersinia* virulence regulon, and AraC, the *Escherichia coli* arabinose operon regulator. *J. Bacteriol.* **171**: 254-262.

Cravioto, A., Gross R. J., Scotland, S. M. and Rowe, B. (1979). An adhesive factor found in strains of *Escherichia coli* belonging to the traditional infantile enteropathogenic serotypes. *Current. Microbiol.* **3**: 95-99.

Cravioto, A., Scotland, S. M. and Rowe, B. (1982). Hemagglutination Activity and Colonization Factor Antigens I and II in Enterotoxigenic and on-Enterotoxigenic Strains of *Escherichia coli* isolated from Humans. *Infect. Immun.* **36**: 189-197.

Dallas, W. S. and Falkow, S. (1980). Aminoacid sequence homology between cholera toxin and *Escherichia coli* heat labile toxin. *Nature (London)* **288**: 499-501.

Danchin, A. and Ullmann, A. (1980). The Co-ordinate expression of polycistronic operons in bacteria. *Trends. Biochem. Sci.*, **5**: 51-52.

Darfeuille, A., Lafeuille, B., Joly, B. and Cluzel, R. (1983). A new colonization factor antigen (CFA/III) produced by enterotoxigenic *Escherichia coli* O128:B12. *Annu. Microbiol.* Paris **134**A: 53-64.

Darfeuille, A., Forestier, C., Joly, B. and Cluzel, R. (1986). Identification of a non-fimbrial adhesive factor of an enterotoxigenic *Escherichia coli* strain. *Infect. Immun.* **52**: 468-475.

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

Dean, A. G., Ching, Y. C., Williams, R. G. and Harden, L. B. (1972). Test for *Escherichia coli* enterotoxin using infant mice: application in a study of diarrhoea in children in Honolulu. *J. Infect. Dis.* **125**: 407-411

De Graaf. F. K. (1990). Genetics of adhesive fimbriae of intestinal *Escherichia coli. Curr. Top. Microbiol. Immunol.* **151**: 29-53.

De Graaf, F. K., Wientjes, F. B. and Klaasen, P. (1980). Production of K99 antigen by Enterotoxigenic *Escherichia coli* strains of antigenic groups O8, O9, O20 and O101 grown at different conditions. *Infect. Immun.* **27**: 216-221.

De Graaf, F. K. and Roorda, I. (1982). Production, purification, and characterisation of the fimbrial adhesive antigen F41 isolated from calf Enteropathogenic *Escherichia coli* strain B41M. *Infect. Immun.* **36**: 751-758.

De Graaf, F. K., Krenn, B. E. and Klaasen, P. (1984). Organisation and expression of genes involved in the biosynthesis of K99 fimbriae. *Infect. Immun.* **43**: 508-514.

De Graaf, F. K. and Klaasen, P. (1986). Organization and expression of genes involved in the biosynthesis of 987P fimbriae. *Molec. Gen. Genet.* **204**: 75-81.

De Graaf, F. K. and Mooi, F. R. (1986). The fimbrial adhesins of *Escherichia coli. Adv. Microb. Physiol.* **28**: 65-143.

De Haan, L. A. M., Willshaw, G. A., van der Zeijst, B. A. M. and Gaastra, W. (1991). The nucleotide sequence of a regulatory gene present in an enterotoxigenic *Escherichia coli* strain of serotype O167: H5. *FEMS Microbiol. Lett.* **83**: 341-346.

Deneke, C. F., Throne, G. M. and Gorbach, S. L. (1981). Serotypes of attachment pili of enterotoxigenic *Escherichia coli* isolated from humans. *Infect. Immun.* **32**: 1254-1260.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the vax. *Nucleic Acids. Res.* **12**: 387-395.

De Wolf, M. J. S., Fridkin, M., Epstein, M. and Kohn, L. D. (1981). Structure-function studies of cholera toxin and its A and B promoters. *J. Biol. Chem.* **256**: 5481-5488.

Diaz-Aroca, E., Mendiola, V. M., Zabala, J. C. and De laCruz, F. (1987). Transposition of IS*91* does not generate a target duplication. *J. Bacteriol.* **169**: 442-443.

Dorman, C. J. and Higgins, C. F. (1987). Fimbrial phase variation in *Escherichia coli*: dependence on Integration Host Factor and homologies with other site-specific recombinases. *J. Bacteriol.* **169**: 3840-3843.

Dougan, G. and Sherratt, D. (1977). The transposon Tn*1* as a probe for studying ColE1 structure and function. *Mol. Gen. Genet.* **151**: 151-160.

Dougan, G., Dowd, G. and Kehoe, M. (1983). Organisation of K88ac encoded polypeptides in the *Escherichia coli* cell envelope: use of minicells and outer membrane protein mutants for studying assembly of pili. *J. Bacteriol.* **153**: 364-370.

Dreyfus, L. A., Frantz, J. C. and Robertson, D. C. (1983). Chemical properties of heat-stable enterotoxins produced by enterotoxigenic *Escherichia coli* of different host origins. *Infect. Immun.* **42**: 537-543.

Dreyfus, L. A., Urban, R. G., Kupersztoch, Y. M., Tachias, K., Slaughter, C. and Whipp, S. C. (1991). Structure and function analysis of the ST_B heat-stable enterotoxin of *Escherichia coli*. In *Twenty-seventh joint conference on cholera and related diarrhoeal diseases*: p111-114.

Duguid, J. P., Smith, I. W., Dempster, G. and Edmunds, P. N. (1955). Non-flagellar filamentous appendages (fimbriae) and haemagglutinating activity in Bacterium coli. *J. Pathol. Bacteriol.* **70**: 335-348.

Duguid, J. P. and Old, D. C. (1980). Adhesive properties of Enterobacteriaeceae. *Bacterial adherence, receptors and recognition, series B,* vol.6. Chapman and Hall, London. pp. 185-217.

Echeverria, P., Seriwatana, J., Taylor, D. N., Changchawalit, S., Smyth, C. J., Twohig, J. and Rowe, B. (1986). Plasmids coding for colonization factor antigens I and II, LT, and ST-A2 in *Escherichia coli*. *Infect. Immun.* **51**: 626-630.

Eisenstein, B. I. (1981). Phase variation of Type I fimbriae in *Escherichia coli* is under transcriptional control. *Science* **214**: 337-339.

Eisenstein, B. I. (1988). Type I fimbriae of *Escherichia coli*: genetic regulation, morphogenesis, and role in pathogenesis. *Rev. Infect. Dis.* **10** (suppl.2): 341-344.

Eisenstein, B. I. and Dodd, D. C. (1982). Pseudocatabolite repression of Type I fimbriae of *Escherichia coli*. *J. Bacteriol.* **151**: 1560-1567.

Elleman, T. C. and Hoyne, P. A. (1984). Nucleotide sequence of the gene encoding pilin of *Bacteriodes nodosus*, the causal organism of bovine footrot. *J. Bacteriol.* **160**: 1184-1187.

Eshdat, Y., Silverblatt, F. J. and Sharon, N. (1981). Dissociation and reassembly of *Escherichia coli* type I pili. *J. Bacteriol.* **148**: 308-314.

Evans, D. G., Silver, R. P., Evans, D. J. Jr., Chase, D. G. and Gorbach, S. L. (1975). Plasmid-controlled colonization factor associated with virulence in *Escherichia coli* enterotoxigenic for humans. *Infect. Immun.* **12**: 656-667.

Evans, D. G. and Evans, D. J. Jr. (1978). New surface associated heat-labile colonization factor antigen (CFA/II) produced by enterotoxigenic *Escherichia coli* of serogroups O6 and O8. *Infect. Immun.* **21**: 638-647.

Fader, R. C., Duffy, L. K., Davis, C. P. and Kurosky, A. (1982). Purification and chemical characterisation of TypeI pili isolated from *Klebsiella pneumoniae*. *J. Biol. Chem.* **257**: 3301-3305.

Falconi, M., Gualtieri, M. T., LaTeana, A., Losso, M. A. and Pon, C. L. (1988). Proteins from the prokaryotic nucleoid: primary and quaternary structure of the 15 kDa *Escherichia coli* DNA binding protein H-NS. *Mol. Microbiol.* **2**: 323-329.

Faris, A., Sellei, J., Lindahl, M. and Wadstrom, T. (1982). *Zentralbl. Bakteriol. Hyg. Abt.* **253**: 175-182.

Field, M., Graf, L. H. Jr., Laird, W. J. and Smith, P. L. (1978). Heat-stable enterotoxin of *Escherichia coli*: *In vitro* effects on guanylate cyclase, cGMP concentration and ion transport in small intestine. *Proc. Natl. Acad. Sci. USA* **75**: 2800-2804.

Files, J. G., Weber, K. and Miller, J. H. (1974). Translational reinitiation: initiation of *lac* represser fragments at three internal sites early in the *lac i* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **71**: 667-670.

Finkelstein, R. A., Burks, M. F., Zupan, A., Dallas, W. S., Jacob, C. O. and Ludwig, D. S. (1987). Antigenic determinants of the cholera/*E. coli* family of enterotoxins. *Rev. Infect. Dis.* suppl.**9**: S490-S502.

Finley, B. B. and Falkow, S. (1989). Common themes in microbial pathogenicity. *Microbiol. Rev.*, **53**: 210-230.

Forsman, K., Göransson, M. and Uhlin, B. E. (1989). Autoregulation and multiple DNA interactions by transcriptional regulatory protein in *Escherichia coli* pili biogenesis. *EMBO. J.* **8**: 1271-1277.

Friedman, D. I. (1988). Integration Host Factor: A protein for all reasons. *Cell* **55**: 545-554.

Fukuta, S., Magnani, J. L., Twiddy, E. M., Holmes, R. K. and Ginsburg, V. (1988). Comparison of the carbohydrate-binding specificities of cholera toxin and *Escherichia coli* heat-labile enterotoxins LTh-I, LT-IIa and LT-IIb. *Infect. Immun.* **56**: 1748-1753.

Gaastra, W. and De Graaf, F. K. (1982). Host-specific fimbrial adhesins of non-invasive enterotoxigenic *Escherichia coli* strains. *Microbiol. Rev.* **46**: 129-161.

Gaastra, W. and Amstrup-Pedersen. (1985). Serological variants of the K88 antigen. *In protein-carbohydrate interactions in biological systems*. Lark, D.L. (ed). London: Academic press, inc. (London) Ltd., pp.95-102.

Gaastra, W., Klemm, P., Walker, J. M. and De Graaf, F. K. (1979). K88 fimbrial proteins: amino and carboxyl terminal sequences of intact proteins and cyanogen bromide fragments. *FEMS Microbiol. Lett.* **6**: 15-18.

Gaastra, W., Mooi, F. R., Stuitje, A. R. and De Graaf, F. K. (1981). The nucleotide sequence of the gene encoding the K88ab protein subunit of porcine enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **12**: 41-46.

Gaastra, W., Klemm, P. and De Graaf, F. K. (1983). The nucleotide sequence of the K88ad protein subunit of porcine enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett,* **18**: 177-183.

Gaastra, W., Jordi, B. J. A. M., Mul, E. M. A., Hamers, A. M., McConnell, M. M., Willshaw, G. A., Smith, H. R. and Van Der Zeijst, B. A. M. (1990). A silent regulatory gene cfaD on region I of the CFA/I plasmid NTP113 of enterotoxigenic *Escherichia coli*. *Microb. Pathog*. **9**: 285-291.

Gallucci, E. and Garen, A. (1966). Suppressor genes for nonsense mutations. II. The *Su-4* and *Su-5* suppressor genes of *Escherichia coli*. *J. Mol. Biol.*, **15**: 193-200.

Garger, S. J., Griffin, O. M. and Grill, L. K. (1983). Rapid purification of plasmid DNA by a single centrifugation in a two step caesium chloride-ethidium bromide gradient. *Biochem. Biophys. Res. Commun.* **117**: 835-842.

Gerlach, G. F., Clegg, S., Ness, N. J., Swenson, D. L., Allen, B. L. and Nichols, W. D. (1989). Expression of type I fimbriae and mannose sensitive hemagglutination by recombinant plasmids. *Infect. Immun.* **57**: 764-770.

Ghosal, D., Sommer, H. and Saedler, H. (1979). Nucleotide sequence of the transposable DNA element IS2. *Nucleic. Acids. Res.* **6**: 1111-1122.

Gill, D. M., Clements, J. D., Robertson, D. C. and Finkelstein, R. A. (1981). Subunit number and arrangement in *Escherichia coli* heat-labile enterotoxin. *Infect. Immun.* **33**: 677-682.

Girardeau, J. P., Dubourguier, H. C. and Contrepois, M. (1980). Attachment des *E. coli* enteropathogenes a la muqueeuse intestinale. *Bull. Group. Technol. Vet.* **80-4-B-190**: 49-59.

Girardeau, J. P., Der Vartanian, M., Ollier, J. L. and Contrepois, M. (1988). CS31A, a new K88-related fimbrial antigen on bovine enterotoxigenic and septicaemic *Escherichia coli* strains. *Infect. Immun.* **56**: 2180-2188.

Girardeau, J. P., Bertin, Y., Martin, Christin., Der Vartanian, M. and Boeuf, C. (1991). Sequence analysis of the *clpG* gene, which encodes for surface antigen CS31A subunit: evidence of an evolutionary relationship between CS31A, K88 and F41 subunit genes. *J. Bacteriol.* **173**: 7673-7683.

Goelet, P., Lomonossoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. and Karn, J. (1982). Nucleotide sequence of tobacco Mosaic virus RNA. *Proc. Natl. Acad. Sci. USA* **79**: 5818-5822.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. and Stormo, G. (1981). Translational initiation in prokaryotes. *Annu. Rev. Microbiol.* **35**: 365-403.

Gold, L. and Stormo, G. (1987). Translation initiation. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. Vol.**2** (ed: Neidhardt, F.C). American Society for Microbiology, Washington, D.C. pp.1302-1307.

Göransson, M, and Uhlin, B. E. (1984). Environmental temperature regulates transcription of a virulence pili operon in *E. coli. EMBO. J.* **3**:2885-2888.

Göransson, M., Forsman, K. and Uhlin, B. E. (1988). Functional and structural homology among regulatory cistrons of pili-adhesin determinants in *Escherichia coli. Mol. Gen. Genet.* **212**: 412-417.

Göransson, M., Forsman, K., Nilsson, P. and Uhlin, B. E. (1989). Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus adhesin in *Escherichia coli. Mol. Microbiol.* **3**: 1557-1565.

Göransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K. and Uhlin, B. E. (1990). Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli. Nature (London)* **344**: 682-685.

Gorbach, S. L., Kean, B. H., Evans, D. G., Evans, D. J. and Bessudo, D. (1975). Traveller's diarrhoea and toxigenic *Escherichia coli. N. Engl. J. Med.* **292**: 933-936.

Gough, J. A. and Murray, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* **166**: 1-19.

Gray, L., Mackman, N., Nicaud, J. -M. and Holland, B. I. (1986). The carboxy-terminal region of haemolysin 2001 is required for secretion of the toxin from *Escherichia coli. Mol. Gen. genet.* **205**: 127-133.

Griffiths, S. L., Finkelstein, R. A. and Critchley, D. R. (1986). Characterisation of the receptor for cholera toxin and *Escherichia coli* heat-labile toxin in rabbit intestinal brush borders. *Biochem. J.* **238**: 313-322.

Gross, R. J., Cravioto, A., Scotland, S. M., Cheasty, T. and Rowe, B. (1978). The occurrence of colonization factor (CF) in enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **3**: 231-233.

Guinee, P. A. M. and Jansen, W. H. (1979). Behaviour of *Escherichia coli* K antigens K88ab, K88ac, and K88ad in immunoelectrophoresis, double diffusion, and hemagglutination. *Infect. Immun.* **23**: 700-705.

Guinee, P. A. M., Agtenberg, C. M., Jansen, W. H. and Frik, J. F. (1977). Serological identification of pig enterotoxigenic *Escherichia coli* strains not belonging to the classical serotypes. *Infect. Immun.* **15**: 549-555.

Guerrant, R. L., Hughes, J. M., Chang, B., Robertson, D. C. and Murad, F. (1980). Activation of intestinal guanylate cyclase by heat-stable enterotoxin of *Escherichia coli*: studies of tissue specificity, potential receptors, and intermediates. *J. Infect. Dis.* **142**: 220-227.

Guth, B. E., Pickett, C. L., Twiddy, E. M., Holmes, R. K., Gomes, T. A., Lima, A. A., Guerrant, R. L., Franco, B. D. and Trabulsi, L. R. (1986a). Production of type II heat-labile enterotoxin by *Escherichia coli* isolated from food and human faeces. *Infect. Immun.* **54**: 587-589.

Guth, B. E., Twiddy, E. M., Trabulsi, L. R. and Holmes, R. K. (1986b). Variation in chemical properties and antigenic determinants among type II heat-labile enterotoxins of *Escherichia coli*. *Infect. Immun.* **54**: 529-536.

Gyles, C. L. (1971). Heat labile and heat stable forms of the enterotoxin from *Escherichia coli* strains enteropathogenic for pigs. *Annu. N. Y. Acad. sci.* **176**: 314-322.

Gyles, C. L., So, M. and Falkow, S. (1974). The enterotoxin plasmids of *Escherichia coli*. *J. Infect. Dis.* **130**: 40-49.

Haenni, A. L., Morch, M. -D., Drugeon, G., Valle, R., Joshi, R. L. and Denial, T. -M. (1987). Gene expression in turnip yellow mosaic virus. *In* M.A. Brinton, and R.R. Rueckert (ed.). *Positive strand RNA viruses*. Alan R. Liss, Inc., New York. pp.149-158.

Hagberg, L., Jodal, U., Korhonen, T. K., Lidin-Janson, G., Lindberg, U. and Svanborg-Eden, C. (1981). Adhesion, hemagglutination and virulence of *Escherichia coli* causing urinary tract infections. *Infect. Immun.* **40**: 564-570.

Hall, R. H. Jr., Maneval, D. R., Collins, J. H., Theibert, J. L. and Levine, M. M. (1989). Purification and analysis of colonization factor antigen I, coli surface antigen 1, and coli surface antigen 3 fimbriae from enterotoxigenic *Escherichia coli*. *J. Bacteriol.* **171**: 6372-6374.

Hamers, A. M., Pel, H. J., Willshaw, G. A., Kusters, J. G., Bernard, A. M., van der Zeijst. and Gaastra, W. (1989). The nucleotide sequence of the first two genes of the CFA/I fimbrial operon of human enterotoxigenic *Escherichia coli*. *Microb. pathog.* **6**: 297-309.

Hanson, M. S. and Brinton, C. C. Jr. (1988). Identification and characterisation of *E. coli* type I pilus tip adhesion protein. *Nature (London)* **332**: 265-268.

Harbison, S. -A., Davies, J. W. and Wilson, T. M. A. (1985). Expression of high molecular weight polypeptides by carnation mottle virus RNA. *J. Gen. Virol.* **66**: 2597-2604.

Hawley, D. K. and McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic. Acids. Res.* 11: 2237-2255.

Hedegaard, L. and Klemm, P. (1989). TypeI fimbriae of *Escherichia coli* as carriers of heterologous antigenic sequences. *Gene* **85**: 115-124.

von Heijne, G. (1983). Pattern of amino acids near signal sequence cleavage sites. *Eur. J. Biochem.* **133**: 17-21.

von Heijne, G. (1984). How signal sequences maintain cleavage specificity. *J. Mol. Biol.* **173**: 243-251.

von Heijne, G. (1985). Signal sequences. The limits of variation. *J. Mol. Biol.* **184**: 99-105.

von Heijne, G. (1986). Net N-C charge imbalance may be important for signal sequence function in bacteria. *J. Mol. Biol.* **192**: 287-290.

Heuzenroeder, M. W., Neal, B. L., Thomas, C. J., Halter, R. and Manning, P. A. (1989). Characterisation and molecular cloning of the PCF8775 CS5 antigen from an enterotoxigenic *Escherichia coli* O115:H40 isolated in central Australia. *Mol. Microbiol.* **3**: 303-310.

Heuzenroeder, M. W., Elliot, T. R., Thomas, C. J., Halter, R. and Manning, P. A. (1990). A new fimbrial type (PCFO9) on enterotoxigenic *Escherichia coli* O9:H^- LT^+ isolated from a case of infant diarrhoea in central Australia. *FEMS Microbiol. Lett.* **66**: 55-60.

Hershey, J. W. B. (1987). Protein synthesis. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. Vol.**1** (ed: Neidhardt, F.C). American Society for Microbiology, Washington, D.C. pp.613-648.

Hibberd, M. L., McConnell, M. M., Field, A. M. and Rowe, B. (1990). The fimbriae of human enterotoxigenic *Escherichia coli* strain 334 are related to CS5. *J. Gen. Microbiol.* **136**: 2446-2456.

Hicks, D. and Molday, R. S. (1984) pp.203-219 In *Science of Biological Specimen Preparation.* Eds Revel, J.P., Barnara, M. T. and Haggis, G.H. SEM, Chicago.

Higgins, D. G. and Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244

Hoekstra, W. P. M., Felix, H. S., Theilemans, M., Pere, A. and Korhonen, T. K. (1986). Chromosomal location of P fimbrial gene clusters in uropathogenic *Escherichia coli* strains. *Proceedings of the 14th International Congress of Microbiology*, IUMS, Manchester, England.

Holland, I. B., Wang, R., Seror, S.J. and Blight, M. (1989). Haemolysin secretion and other protein translocation mechanisms in Gram-negative bacteria. *Soc. Gen. Microbiol. Symposium* **44**: Baumberg, S., Hunter, I. and Rhodes, M. (eds.) Cambridge, Cambridge University Press, pp. 219-254.

Holmes, R. K., Pickett, C. L., Belisle, B. W., Twiddy, E. M., Newland, J. W., Strockbine, N. A., Miller, S. F. and O'Brien, A. D. (1986a). Genetics of toxinogenesis in *Escherichia coli* relating to vaccine development.. *In* J, Holmgren A, Lindberg, and Mollby, R. (ed.), *Development of vaccines and drugs against diarrhoea.* 11th Nobel conference, Stockholm, 1985. Student literature, Lund, Sweden. pp.68-73

Holmes, R. K., Twiddy, E. M. and Pickett, C. L. (1986b). Purification and characterisation of type II heat-labile enterotoxin of *Escherichia coli. Infect. Immun.*, **53**: 464-473.

Holmes, R. K., Pickett, C. L. and Twiddy, E. M. (1988). Genetic and biochemical studies of type II heat-labile enterotoxins of *Escherichia coli. Zentrabl. Bacteriol. Suppl.* **17**: 187-194.

Holmes, R. K., Twiddy, E. M., Pickett, C. L., Marcus, H., Jobling, M. G. and Petitjean, F. M. J. (1990). The *Escherichia coli / Vibrio cholerae* family of enterotoxins, pp. 91-102. In Pohland, A. E., Dowell, V. R. Jr. and Richard, J. L. (ed. *symposium on molecular mode of action of selected microbial toxins in foods and feeds*) Plenum press, New York.

Honda, T., Arita, M. and Miwatani, M. (1984). Characterisation of new hydrophobic pili of human enterotoxigenic *Escherichia coli*: a possible new colonization factor. *Infect. Immun.* **43**: 959-965.

Honda, T., Cakir, N., Arita, M. and Miwatani, T. (1989). Purification and characterisation of the CS2 pili of colonization factor antigen II produced by human enterotoxigenic *Escherichia coli*. *Microbiol. Immunol.* **33**: 265-275.

Hoopes, B. C. and McClure, W. R. (1987). Strategies in regulation of transcription initiation. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. Vol.**2** (ed: Neidhardt, F.C). American Society for Microbiology, Washington, D.C. pp. 1231-1241.

Houwink, A. L. and Van Itersen, W. (1950). Electron microscopical observations on bacterial cytology. II. A study on flagellation. *Biochem. Biophys. Acta.* **5**:10-44.

Hromockyj, A. E. and Maurelli, A. T. (1989). Identification of an *Escherichia coli* gene homologous to virR, a regulator of *Shigella* virulence. *J. Bacteriol.* **171**: 2879-2881.

Hsu, Y. H. and Brakke, M. K. (1985). Cell-free translation of soil-borne wheat mosaic virus RNAs. *Virology* **143**: 272-279.

Hughes, G., Davies, J. W. and Wood, K. R. (1986). *In vitro* translation of bipartite genomic RNA of pea early browning virus. *J. Gen. Virol.* **67**: 2125-2133.

Hull, R., Bieler, S., Falkow, S. and Hull, S. (1986). Chromosomal map position of genes encoding P-adhesins in uropathogenic *Escherichia coli*. *Infect. Immun.* **51**: 693-695.

Hultgren, S. J., Lindberg, F., Magnusson, G., Kihlberg, J., Tennent, J. M. and Normark, S. (1989). The PapG adhesin of uropathogenic *Escherichia coli* contains separate regions for receptor binding and for pilus incorporation. *Proc. Natl. Acad. Sci. USA* **86**: 4357-4361.

Hultgren, S. J., Duncan, J. I., Schaeffer, A. J. and Amundsen, S. K. (1990). Mannose-sensitive hemagglutination in the absence of piliation in *Escherichia coli*. *Mol. Microbiol.* 4: 1311-1318.

Hultgren, S. J., Abraham, S. N. and Normark, S. (1991). Chaperone-assisted assembly and molecular architecture of adhesive pili. *Annu. Rev. Microbiol.* **45**: 383-415.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**: 13-34.

Isaacson, R. E. (1977). K99 surface antigen of *Escherichia coli*: purification and partial characterisation. *Infect. Immun.* **15**: 272-279.

Isaacson, R. E. (1980a). Pili of enterotoxigenic *Escherichia coli*. *In*: Third international symposium on neonatal diarrhoea. (S Acres ,ed) Veterinary Infectious Disease Organisation, Ssakatoon. pp. 213-236.

Isaacson, R. E. (1980b). Factors affecting expression of the *Escherichia coli* pilus K99. *Infect. Immun.* **28**: 190-194.

Isaacson, R. E., Nagy, B. and Moon, H. W. (1977). Colonization of porcine small intestines by *Escherichia coli*: colonization and adhesion factors of pig enteropathogens that lack K88. *J. Infect. Dis.* **135**: 531-539

Isaacson, R. E. and Richter, P. (1981). Escherichia *coli* 987P pilus: purification and partial characterisation. *J. Bacteriol.* **46**: 784-789.

Isaacson, R. E., Colmenero, J. and Richter, P. (1981). *Escherichia coli* K99 pili are composed of one subunit species. *FEMS Microbiol. Lett.* **12**: 229-232.

Ishikawa, M., Meshi, T., Motoyoshi, F., Takamatsu, N. and Okada, Y. (1986). In vitro mutagenesis of the putative replicase genes of tobacco mosaic virus. *Nucleic. Acids. Res.* **14**: 8291-8305

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**: 318-356.

Jacobs, A. A. C., Venema, J., Leeven, R., Van Pelt-Heerschap, H. and De Graaf, F. K. (1987a). Inhibition of adhesive activity of K88 fimbrillae by peptides derived from the K88 adhesin. *J. Bacteriol.* **169**: 735-741.

Jacobs, A. A. C., Simons, B. H. and De Graaf, F. K. (1987b). The role of lysine-132 and arginine-136 in the receptor-binding domain of the K99 fibrillar subunit. *EMBO. J.* **6**: 1805-1808.

Jann, K., Jann, B. and Schmidt, G. (1981). SDS polyacrylamide gel electrophoresis and serological analysis of pili from *Escherichia coli* of different pathogenic origin. *FEMS Microbiol. Lett.* **11**: 21-25.

Jennings, M. L. (1989). Topography of membrane proteins. *Annu. Rev. Biochem.* **58**: 999-1027.

Jones, G. W. and Isaacson, R. E. (1983). Proteinaceous bacterial adhesins and their receptors. *CRC. Crit. Rev. Microbiol.* **10**: 229-260.

Jovanovich, S. B., Martinell, M., Record, M. T. Jr. and Burgess, R. R. (1988). Rapid response to osmotic upshift by osmoregulated genes in *Escherichia coli* and *Salmonella typhimurium. J. Bacteriol.* **170**: 534-539.

Kahn, M., Kotler, T., Thomas, C., Figurski, D., Meyer, R., Renault, E. and Helinski, D. R. (1979). Plasmid cloning vehicles derived from plasmids ColE1, RGK and RK2. *Methods. Enzymol.* **68**: 268-280.

Kallenius, G., Sevenson, S. B., Hultberg, H., Molby, R., Helin, I., Cedergren, B. and Winberg, J. (1981). Occurrence of P-fimbriated *Escherichia coli* in urinary tract infections. *Lancet* **ii**: 1369-1372.

Kalionis, B., Dodd, I. B. and Egan, J. B. (1986). Control of gene expression in the P2-related temperate coliphages. DNA sequence of the major control region of phage 186. *J. Mol. Biol.* **191**: 199-209.

Karjalainen, T. K., Evans, D. G., So, M. and Lee, C. H. (1989). Molecular cloning and nucleotide sequence of the colonization factor antigen I gene of *Escherichia coli. Infect. Immun.* **57**: 1126-1130.

Karjalainen, T. K., Evans, D. G., Evans, D. J. Jr., Graham, D. Y. and Lee, C. H. (1991). Iron represses the expression of CFA/I fimbriae of enterotoxigenic *E. coli. Microb. Pathog.*, **11**: 317-323.

Karr, J. F., Nowicki, B., Truong, L. D., Hull, R. and Hull, S. (1989). Purified P fimbriae from two cloned gene clusters of a single pyelonephritogenic strain adhere to unique structures in the human kidney. *Infect. Immun.* **57**: 3594-3600.

Karr, J. F., Nowicki, B., Truong, L. D., Hull, R., Moulds, J. and Hull, S. (1990). Pap-2 encoded fimbriae adhere to the P-blood group-related glycosphingolipid Stage-Specific Embryonic Antigen 4 in the human kidney. *Infect. Immun.* **58**: 4055-4062.

Kennedy, N., Beutin, L., Achtman, M., Skurray, R., Rahmsdorf, U. and Herrlich, P. (1977) Conjugation proteins of the F sex factor. *Nature (London)* 270: 580-585.

Klemm, P. (1981). The complete amino acid sequence of the K88 antigen, a fimbrial protein from *Escherichia coli. Eur. J. Biochem.* **117**: 617-627.

Klemm, P. (1982). Primary structure of the CFA/I fimbrial protein from human enterotoxigenic *Escherichia coli* strains. *Eur.J.Biochem.* **124**: 339-348.

Klemm, P. (1984). The fimA gene encoding the TypeI fimbrial subunit of *Escherichia coli:* nucleotide sequence and primary structure of the protein. *Eur. J. Biochem.* **143**: 395-399.

Klemm, P. (1985). Fimbrial adhesins of *Escherichia coli. Rev. Infect .Dis.* 7: 321-340.

Klemm, P. (1986). Two regulatory fim genes fimB and fimE control the phase variation of type I fimbriae in *Escherichia coli. EMBO. J.* **5**: 1389-1393.

Klemm, P. and Christiansen, G. (1987). Three fim genes required for the regulation of length and mediation of adhesion of *Escherichia coli* TypeI fimbriae. *Mol. Gen. Genet.* **208**: 439-445.

Klemm, P. and Christiansen, G. (1990). The fimD gene required for cell surface localisation of *Escherichia coli* TypeI fimbriae. *Mol. Gen. Genet.* **220**: 334-338.

Klemm, P. and Mikkelsen, L. (1982). Prediction of antigenic determinants and secondary structures of the K88 and CFA/I fimbrial proteins from enteropathogenic *Escherichia coli*. *Infect. Immun.* **38**: 41-45.

Klemm, P., Orskov, I. and Orskov, F. (1982). F7 and type I-like fimbriae from three *Escherichia coli* strains isolated from urinary tract infections: protein, chemical and immunological aspects. *Infect. Immun.* **36**: 462-468.

Klemm, P., Gaastra, W., McConnell, M. M. and Smith, H. R. (1985a). The CS2 fimbrial antigen from *Escherichia coli*, purification, characterisation and partial covalent structure. *FEMS Microbiol. Lett.* **26**: 207-210.

Klemm, P., Jorgensen, B. J., Van Die, I., De Ree, H. and Bergmans, H. (1985b). The *fim* genes responsible for synthesis of type I fimbriae in *Escherichia coli* cloning and genetic organization. *Mol. Gen. Genet.* **99**: 410-414.

Knutton, S., Lloyd, D. R., Candy, D. C. A. and McNeish, A. S. (1984). Ultrastructural study of adhesion of enterotoxigenic *Escherichia coli* to erythrocytes and human intestinal epithelial cells. *Infect. Immun.* **44**: 519-527.

Knutton, S., Lloyd, D. R. and McNeish, A. S. (1987). Identification of a new fimbrial structure in enterotoxigenic *Escherichia coli* (ETEC) serotype O148:H28 which adheres to human intestinal mucosa: a potentially new human ETEC colonization factor. *Infect.Immun.* **55**: 86-92.

Korhonen, T. K., Vaisanen, V., Saxen, H., Hultberg, H. and Svenson, S. B. (1982). P-antigen recognising fimbriae from human uropathogenic *Escherichia coli* strains. *Infect.Immun.* **37**: 286-291.

Korhonen, T. K., Vaisanen-Rhen, V., Rhen, M., Pere, A., Parkkinen, J. and Finne, J. (1984). *Escherichia coli* fimbriae recognising sialyl galactosides. *J. Bacteriol.* **159**: 762-766.

Korhonen, T. K., Valtonen, M. V., Paarkkinen, J., Vaisanen-Rhen, V., Finne, J., Orskov, F., Orskov, I., Svenson, S. B. and Makela, P. H. (1985). Serotypes, haemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. *Infect. Immun.* **48**: 486-491.

Koronakis, V., Koronakis, E. and Hughes, C. (1989). Isolation and analysis of the C-terminal signal directing export of *Escherichia coli* hemolysin protein across both bacterial membranes. *EMBO J.* **8**: 595-605.

Krallmann-Wetzel, U., Ott, M., Hacker, J. and Schmidt, G. (1989). Chromosomal mapping of genes encoding mannose-sensitive (type I) and mannose-resistant F8 (P) fimbriae of *Escherichia coli* O18:K5:H5. *FEMS Microbiol. Lett.* **58**: B15-B22.

Krogfelt, K. A., Bergmans, H. and Klemm, P. (1990). Direct evidence that the FimH protein is the mannose specific adhesin of *Escherichia coli* type I fimbriae. *Infect.Immun.* **58**: 1995-1999.

Kuehn, M. J., Normark, S. and Hultgren, S. (1991). Immunoglobulin-like PapD chaperone caps and uncaps interactive surfaces of nacently translocated pilus subunits. *Proc. Natl. Acad. Sci. USA* **88**: 10586-10590

Kuehn, M. J., Heuser, J., Normark, S. and Hultgren, S. (1992). P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature (London)* **356**: 252-255.

Kupersztoch, Y. M., Tachias, K., Moomaw, C. R., Dreyfus, L. A., Urban, R. G., Slaughter, C. and Whipp, S. C. (1990). Secretion of methanol insoluble heat-stable enterotoxin (ST_B): energy- and *sec*A- dependent conversion of pre-ST_B to an intermediate indistinguishable from the extracellular toxin. *J. Bacteriol.* **172**: 2427-2432.

Kurland, C. G. (1987). Strategies for efficiency and accuracy in gene expression. 1.The major codon preference: a growth optimisation strategy. *Trends. Biochem. Sci.* **12**: 126-128.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. Appendix:- A C program for evaluating the hydropathic character of sequence segments. *J. Mol. Biol.* **157**: 105-132.

Lark, D. L., Normark, S., Uhlin, B. -E. and Wolf-Watz, H. (1986). Protein-carbohydrate interactions in biological systems. *FEMS symposium no.31.* Academic press, London, pp. 464.

Lee, C. H., Moseley, S. L., Moon, H. W., Whipp, S. C., Gyles, C. L. and So, M. (1983). Characterisation of the gene encoding heat-stable toxin II and preliminary molecular epidemiology studies of enterotoxigenic *Escherichia coli* heat-stable toxin II producers. *Infect. Immun.* **42**: 264-268.

Levine, M. M. (1981). Adhesion and micro-organism pathogenicity. *Ciba Found. Symp.*, **80**: 142-545.

Levine, M. M. (1987). *Escherichia coli* that cause diarrhoea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohaemorrhagic and enteroadherent. *J. Infect. Dis.* **155**: 377-385.

Levine, M. M., Ristanio, P., Marley, G., Smyth, C. J., Knutton, S., Boedeker, E., Black, R., Young, C., Clements, M., Cheney, C. and Patnaik, R. (1984). Coli Surface Antigens 1 and 3 of Colonization Factor Antigen II positive Enterotoxigenic *Escherichia coli:* Morphology, Purification, and Immune responses in Humans. *Infect. Immun.* **44**: 409-420.

Lindahl, M., Faris, A., Wadstrom, T. and Hjerton, S. (1981). A new test based on "salting out" to measure relative surface hydrophobicity of bacterial cells. *Biochem. Biophys. Acta.* **677**: 471-476

Lindberg, F., Lund, B. and Normark, S. (1984). Genes of Pyelonephritogenic *E. coli* required for digalactoside-specific agglutination of human cells. *EMBO. J.* **3**: 1167-1173.

Lindberg, F., Lund, B. and Normark, S. (1986). Gene products specifying adhesion of uropathogenic *Escherichia coli* are minor components of pili. *Proc. Natl. Acad. Sci. USA* **83**: 1891-1895.

Lindberg, F., Lund, B., Johansson, L. and Normark, S. (1987). Localisation of the receptor-binding protein at the tip of the bacterial pilus. *Nature (London)* **328**: 84-87.

Lindberg, F., Tennent, J. M., Hultgren, S. J., Lund, B. and Normark, S. (1989). PapD, a periplasmic transport protein in P-pilus biogenesis. *J. Bacteriol.* **171**: 6052-6058.

Lintermans, P., Pohl, P., Deboeck, F., Bertels, A., Schlicker, J., Vandekerckhove, J., Van Damme, J., Van Montagu, M. and De Greve, H. (1988). Isolation and nucleotide sequence of the F17-A gene encoding the structural protein of the F17 fimbriae in bovine enterotoxigenic *Escherichia coli*. *Infect. Immun.* **56**: 1475-1484.

Lintermans, P. F., Bertels, A., Schlicker, C., Deboeck, F., Charlier, G., Pohl, P., Norgren, M., Normark, S., Van Montagu, M. and De Greve, H. (1991). Identification, characterisation, and nucleotide sequence of the F17-G gene, which determines receptor binding of *Escherichia coli* F17 fimbriae. *J. Bacteriol.* **173**: 3366-3373.

Lipman. D. J. and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**: 1435-1440.

Low, D. A., Robinson, E. N. Jr., McGee, Z. A. and Falkow, S. (1987). The frequency of expression of pyelonephritis-associated pili is under regulatory control. *Mol. Microbiol.* **1**: 335-346.

Ludwig, D. S., Holmes, R. K. and Schoolnik, G. K. (1985). Chemical and immunochemical studies on the receptor binding domain of cholera toxin B subunit. *J. Biol. Chem.* **260**: 12528-12534.

Lugtenberg, B., Meijers, J., Peters, R., van der Hoek, P. and van Alphen, L. (1975). Electrophoretic resolution of the major outer membrane protein of *Escherichia coli* K-12 into four major bands. *FEBS Lett.* **58**: 254-258.

Lund, B., Lindberg, F. P., Båga, M. and Normark, S. (1985). Globoside-specific adhesins of uropathogenic *Escherichia coli* are encoded by similar transcomplementable gene clusters. *J. Bacteriol.* **162**: 1293-1301.

Lund, B., Lindberg, F., Marklund, B. -I. and Normark, S. (1987). The PapG protein is the α-D-galactopyranosyl-(1-4) β-D-Galactopyranose-binding adhesin of uropathogenic *Escherichia coli. Proc. Natl. Acad. Sci. USA* **84**: 5898-5902.

Lund, B., Marklund, B. -I., Stromberg, N., Lindberg, F., Karlsson, K. -A. and Normark, S. (1988). Uropathogenic *Escherichia coli* express serologically identical pili with different receptor binding specificities. *Mol. Microbiol.* **2**: 255-263.

Mackman, N., Nicaud, J. -M., Gray, L. and Holland, I.B. (1985). Identification of polypeptides required for the export of haemolysin 2001 from *E. coli. Mol. Gen. Genet.* **201**: 529-536.

Mackman, N., Nicaud, J. -M., Gray, L. and Holland, I.B. (1986). Secretion of haemolysin by *Escherichia coli. Curr. Topi. Microbiol. Immun.* **125**: 159-181.

MacRae, T. H., Dobson, W. J. and McCurdy, H. D. (1977). Fimbriation in gliding bacteria. *Can. J. Microbiol.* **23**: 1096-1108.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular cloning. A laboratory manual.* Cold Spring Harbor, New york: Cold Spring Harbor Laboratory Press.

Manley, J. L. (1978). Synthesis and degradation of termination and premature-termination fragments of β-galactosidase *in vitro* and *in vivo*. *J. Mol. Biol.* **125**: 407-432.

Manning, P. A. and Reeves, P. (1977). Outer membrane protein 3B of *Escherichia coli* K-12 : effect of growth temperature on the amounts of the protein and further characterization on acrylamide gels. *FEMS Microbiol. Lett.* **1**: 275-278.

Manning, P. A., Timmis, K. N. and Stevenson, G. (1985). Colonization Factor Antigen II (CFA/II) of Enterotoxigenic *Escherichia coli:* molecular cloning of the CS3 determinant. *Mol. Gen. Genet.* **200**: 322-327.

Manning, P. A., Higgins, G. D., Lumb, R. and Lanser, J. A. (1986). Colonization factor antigens and a new fimbrial type, CFA/V on O115:H40 and H^- strains of enterotoxigenic *Escherichia coli* in central Australia. *J. Infect. Dis.* **156**: 841-844.

Manoil, C. and Beckwith, J. R. (1985) Tn*phoA*: a transposon probe for protein export signals. *Proc. Natl. Acad. Sci. USA* **82**: 8129-8133.

Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986). Codon usage tabulated from the Genbank genetic sequence data. *Nucleic. Acids. Res.***14 (23)**: r151.

Maurer, L. and Orndorff, P. E. (1985). A new locus, *pilE*, required for the binding of type I piliated *Escherichia coli* to erythrocytes. *FEMS Microbiol. Lett.* **30**: 59-66.

Maurer, L. and Orndorff, P. E. (1987). Identification and characterisation of genes determining receptor binding and pilus length of *Escherichia coli* TypeI pili. *J. Bacteriol.* **169**: 640-645.

McConnell, M. M., Thomas, L. V., Willshaw, G. A., Smith, H. R. and Rowe, B. (1988). Genetic control and properties of coli surface antigens of colonization factor antigen IV (PCF8775) of enterotoxigenic *Escherichia coli. Infect. Immun.* **56**: 1974-1980.

McConnell, M. M., Chart, H., Field, A. M., Hibberd, M. and Rowe, B. (1989). Characterisation of a putative colonization factor (PCFO166) of enterotoxigenic *Escherichia coli* of serogroup O166. *J. Gen. Microbiol.* **135**: 1135-44.

McConnell, M. M., Hibberd, M., Field, A. M., Chart, H. and Rowe, B. (1990). Characterisation of a new putative colonization factor (CS17) from a human enterotoxigenic *Escherichia coli* of serotype O114:H21 which produces only heat-labile enterotoxin. *J. Infect. Dis.* **161**: 343-347.

McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. (1981). A system to study promoter and terminator signals recognised by *Escherichia coli* RNA polymerase. In Chirikijian, J.G. and Papas, T.S. (eds), Gene amplification and analysis, Vol.II: structural analysis of Nucleic Acids. Elsevier, New York, 1981, pp. 383-415.

McMichael, J. C. and Ou, J. T. (1979). Structure of common pili from *Escherichia coli*. *J. Bacteriol.* **138**: 969-975.

Mekalanos, J. J. (1985). Cholera toxin: genetic analysis, regulation, and role in pathogenesis. *Curr. Top. Microbiol. Immun.* **118**: 97-118.

Merson, M. H., Morris, G. K., Sack, D. A., Wells, J. G., Feeley, J. C., Sack, R. B., Creech, W. B. and Kapikian, A. Z. (1976). Traveller's diarrhoea in Mexico: a prospective study of physicians and family members attending a congress. *N. Engl. J. Med.* **294**: 1299-1205.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either strand and double-digest restriction fragments. *Gene* **19**: 269-276.

Middlebrook, J. L. and Dorland, R. B. (1984). Bacterial toxins: cellular mechanisms of action. *Microbiol. Rev.* **48**: 199-221.

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

Moch, T., Hoschutzky, H., Hacker, J., Kronke, K. D. and Jann, K. (1987). Isolation and characterisation of the α- sialyi-β-2-3-galactosyl-specific adhesin from fimbriated *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84**: 3462-3466.

Monckton, R. P. and Hasse, D. (1988). Detection of enterotoxigenic *Escherichia coli* in piggeries in Victoria by DNA hybridisation using K88, LT, ST-1 and ST-2 probes. *Vet. Microbiol.***16**: 273-281.

Mooi, F. R. and De Graaf, F. K. (1979). Isolation and characterisation of K88 antigens. *FEMS Microbiol. Lett.* **5**: 17-20.

Mooi, F. R., De Graaf, F. K. and Van Embden, J. D. A. (1979). Cloning , mapping, and expression of the genetic determinant that encodes for the K88ab antigen. *Nucleic. Acids. Res.*, **6**: 849-865.

Mooi, F. R. and De Graaf, F. K. (1985). Molecular biology of fimbriae of enterotoxigenic *Escherichia coli*. *Curr. Top. Microbiol. Immun.* **118**: 119-138.

Mooi, F. R., Harms, N., Bakker, D. and De Graaf, F. K. (1981). Organisation and expression of genes involved in the production of the K88ab antigen. *Infect. Immun.* **32**: 1155-1163.

Mooi, F. R., Wouters, C., Wijfjes, A. and De Graaf, F. K. (1982). Construction and characterisation of mutants impaired in the biosynthesis of the K88ab antigen. *J. Bacteriol.* **150**: 512-521.

Mooi, F. R., Wijfjes, A. and De Graaf, F. K. (1983). Identification and characterisation of precursors in the biosynthesis of the K88ab fimbriae of *Escherichia coli*. *J. Bacteriol.* **154**: 41-49.

Mooi, F. R., van Buuren, M., Koopman, G., Roosendaal, B. and De Graaf, F. K. (1984). K88ab gene of *Escherichia coli* encodes a fimbria-like protein distinct from the K88ab fimbrial adhesin. *J. Bacteriol.* **159**: 482-487.

Mooi, F. R., Claasen, I., Bakker, D., Kuipers, H. and De Graaf, F. K. (1986). Regulation and structure of an *Escherichia coli* gene coding for outer membrane

protein involved in the export of the K88ab fimbrial subunits. *Nucleic. Acids. Res.* **14**: 2443-2457.

Moon, H. W., Nagy, B., Isaacson, R. E. and Orskov, I. (1977). Occurrence of K99 antigen on *Escherichia coli* isolated from pigs and colonisation of pig ileum by K99$^+$ enterotoxigenic *E. coli* from calves and pigs. *Infect. Immun.* **15**: 614-620.

Moon, H. W., Schneider, R. A. and Moseley, S. L. (1986). Comparative prevalence of four enterotoxin genes among *Escherichia coli* isolated from swine. *Am. J. Vet. Res.* **47**: 210-212.

Moseley, S. L., Echeverria, P., Seriwatana, J., Tirapat, C., Chaicumpa, W., Sakuldaipeara, T. and Falkow, S. (1982). Identification of enterotoxigenic *Escherichia coli* by colony hybridisation using three enterotoxin gene probes. *J. Infect. Dis.* **145**: 863-869.

Moseley, S. L., Hardy, J. W., Huq, M. I., Echeverria, P. and Falkow, S. (1983). Isolation and nucleotide sequence determination of a gene encoding a heat-stable enterotoxin of *Escherichia coli*. *Infect. Immun.* **39**: 1167-1174.

Moseley, S. L., Dougan, G., Schneider, R. A. and Moon, H. W. (1986). Cloning of chromosomal DNA encoding the F41 adhesin of enterotoxigenic *Escherichia coli* and genetic homology between adhesins F41 and K88. *J. Bacteriol.* **167**: 799-804.

Mouricout, M. and Julien, R. (1986). Inhibition of mannose-resistant hemagglutination of sheep erythrocytes by enterotoxigenic *Escherichia coli* in the presence of plasma glycoprotein glycans. *FEMS Microbiol. Lett.* **37**: 145-149.

Mullany, P., Field, A. M., McConnell, M. M., Scotland, S. M., Smith, H. R. and Rowe, B. (1983). Expression of plasmids coding for colonization factor antigen II (CFA/II) and enterotoxin production in *Escherichia coli*. *J. Gen. Microbiol.* **129**: 3591-3601.

Nash, H. A. (1981). Integration and excision of bacteriophage l: the mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **15**: 143-167.

Norgren, M., Normark, S., Lark, D., O'Hanley, P., Schoolnik, G., Falkow, S., Svanborg-Eden, C., Båga, M. and Uhlin, B. E. (1984). Mutations in *Escherichia coli* cistrons affecting adhesion to human cells do not abolish Pap pili fiber formation. *EMBO. J.* **3**: 1159-1165.

Norgren, M., Båga, M., Tennent, J. M. and Normark, S. (1987). Nucleotide sequence, regulation and functional analysis of the PapC gene required for cell surface localisation of pap pili of uropathogenic *Escherichia coli*. *Mol. Microbiol.*, **1**: 169-179.

Normark, S., Lark, D., Hull, R., Norgren, M., Båga, M., O' Hanley, P., Schoolnik, G. and Falkow, S. (1983a). Genetics of digalactoside-binding adhesin from a uropathogenic *Escherichia coli* strain. *Infect. Immun.* **41**: 942-949.

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

Normark, S., Båga, M., Göransson, M., Lindberg, F. P., Lund, B., Norgren, M. and Uhlin, B. E. (1986). pp. 113-143 *In* Mirelman, D. (ed.) *Microbial Lectins and agglutinins*: Properties and biological activity. John Wiley and sons, New York.

Nowicki, B., Moulds, J., Hull, R. A. and Hull, S. I. (1988). A hemagglutinin of uropathogenic *Escherichia coli* recognises the Dr blood group antigen. *Infect. Immun.* **56**: 1057-1060.

O'Hanley, P., Lark, D., Falkow, S. and Schoolnik, G. (1985). Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice. *J. Clin. Invest.* **75**: 347-360.

Old, D. C. (1972). Inhibition of interaction between fimbrial haemagglutinins and erythrocytes by D-mannose and other carbohydrates. *J. Gen. Microbiol.* **71**: 149-157.

Ollier, J. L. and Girardeau, J. P. (1983). Structures et fontions hemagglutinantes des pili K99 dont la biosynthese est glucose-dependante ou constitutive et des pili F41. *Annu. Microbiol.* **134A**: 247-254.

Olsen, A., Jonsson, A. and Normark, S. (1989). Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*. *Nature (London)* **338**: 652-655.

Olsson, E. and Soderlind, O. (1980). Comparison of different assays for definition of heat-stable enterotoxigenicity of *Escherichia col*i porcine strains. *J. Clin. Micribiol.* **11**: 6-15.

Orndorff, P. E. and Falkow, S. (1984). Organization and expression of genes responsible for type I piliation in *Escherichia coli*. *J. Bacteriol.* **159**: 736-744.

Orndorff, P. E. and Falkow, S. (1985). Nucleic acid sequence of pilA, the gene encoding the structural component of TypeI pili in *Escherichia coli*. *J. Bacteriol.* **162**: 454-457.

Oro, H. S., Kolsto, A. B., Wenneras, C. and Svennerholm, A. M. (1990). Identification of asialo GM1 as a binding structure for *Escherichia coli* colonization factor antigens. *FEMS Microbiol. Lett.* **60**: 289-292.

Orskov, I. and Orskov, F. (1966). Episome-carried surface antigen K88 of *Escherichia coli*. 1. Transmission of the K88 antigen and influence on the transfer of chromosomal markers. *J. Bacteriol.* **91**: 69-75.

Orskov, I. and Orskov, F. (1983). Serology of *Escherichia coli* fimbriae. *Prog. Allergy*. **33**: 80-105.

Orskov, I. and Orskov, F. (1985). *Escherichia coli* in extra intestinal infections. *J.Hyg.*, Camb. **95**: 551-575.

Orskov, I., Orskov, F., Sojka, W. J. and Wittig, W. (1964). K. antigens K88ab(L) and K88ac(L) in *Escherichia coli*. A new O antigen: O147 and a new K89(B). *Acta. Pathol. Microbiol. Scand. Sect. B*. **62**: 439-447.

Orskov, I., Orskov, F., Smith, H. W. and Sojka, W. J. (1975). The establishment of K99, a thermolabile, transmissible *Escherichia coli* K antigen, previously called "Kco" possessed by calf and lamb enterotoxigenic strains. *Acta. Pathol. Microbiol. Scandi. Sect. B.* **83**: 31-36.

Orskov, I., Orskov, F., Jann, B. and Jann, K. (1977). Serology, chemistry and genetics of O and K antigens of *Escherichia coli*. *Microbial Rev.* **41**: 667-710.

Orskov, I., Orskov, F. and Birch-Andersen, A. (1980). Comparison of *Escherichia coli* fimbrial antigen F7 with type I fimbriae. *Infect. Immun.* **27**: 657-666.

Osborn, M. J., Gander, J. E., Parisi, E. and Carson, J. (1972a). Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. *J. Biol. Chem.* **247**: 3962-3971.

Osborn, M. J., Gander, J. E. and Parisi, E. (1972b). Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. Site of synthesis of lipopolysaccharides. *J. Biol. Chem.* **247**: 3973-3985.

Ottow, J. C. G. (1975). Ecology, Physiology, and genetics of fimbriae and pili. *Annu. Rev. Microbiol.* **29**: 79-108.

Oudega, B. and De Graaf, F. K. (1988). Genetic organisation and biogenesis of adhesive fimbriae of *Escherichia coli*. *Antonie. van. Leeuwenhoek.* **54**: 285-299.

Oudega, B., De Graaf, M., De Boer, L., Bakker, D., Vader, C. E. M., Mooi, F. R. and De Graaf, F. K. (1989). Detection and identification of faeC as a minor component of K88 fibrillae of *Escherichia coli*. *Mol. Microbiol.* **3**: 645-652.

Pallesen, L., Madsen, O. and Klemm, P. (1989). Regulation of the phase switch controlling expression of type I fimbriae in *Escherichia coli*. *Mol. Microbiol.* **3**: 925-931.

Paranchych, W. and Frost, L. S. (1988). The physiology and biochemistry of pili. *Adv. Microb. Physiol.* **29**: 53-114.

Parker, J. (1989). Errors and alternatives in reading the universal genetic code. *Microbiol. Rev.* **53**: 273-298.

Paterson, R. and Knight, C. A. (1975). Protein synthesis in tobacco protoplasts infected with tobacco mosaic virus. *Virology* **64**: 10-22.

Pearce, W. A. and Buchanan, T. M. (1980). Structure and cell membrane-binding properties of bacterial fimbriae. *In Bacterial adherence,* Beachey, E.H., ed., Chapman and Hall, London. pp. 289.

Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**: 2444-2448.

Pedersen, P. A. (1991). Structure-function analysis of the K88ab fimbrial subunit protein from porcine enterotoxigenic *Escherichia coli. Mol. Microbiol.* **5**: 1073-1080.

Pelham, H. R. B. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature (London)* **272**: 469-471.

Penaranda, M. E., Evans, D. G., Murray, B. E. and Evans, D. J. Jr. (1983). Plasmids in enterotoxigenic *Escherichia coli* belonging to serogroups O6, O8, O85 and O139. *J. Bacteriol.* **154**: 980-983.

Perez-Casal, J., Swartley, J. S. and Scott, J. R. (1990). Gene encoding the major subunit of CS1 pili of human enterotoxigenic *Escherichia coli. Infect. Immun.* **58**: 3594-3600.

Picken, R. N., Mazaitis, A. J., Maas, W. K., Rey, M. and Heynecker, H. (1983). Nucleotide sequence of the gene for heat-stable enterotoxin II of *Escherichia coli. Infect. Immun.* **42**: 269-275.

Pickett, C. L. and Holmes, R. K. (1990). Nucleotide sequence of *Escherichia coli* heat-labile toxins type IIa and IIb and comparisons to type I enterotoxin and cholera toxin, pp. 167-171. *In* Sack, R.B., Zinnaka, Y (ed.), *Advances in research on cholera and related diarrhoeas,* Vol.7. KTK Scientific publishers, Tokyo.

Pickett, C. L., Twiddy, E. M., Belisle, B. W. and Holmes, R. K. (1986). Cloning of genes that encode a new heat-labile enterotoxin of *Escherichia coli*. *J. Bacteriol.* **165**: 348-352.

Pickett, C. L., Weinstein, D. L. and Holmes, R. K. (1987). Genetics of type IIa heat-labile enterotoxin of *Escherichia coli:* operon fusions, nucleotide sequence, and hybridisation studies. *J. Bacteriol.* **169**: 5180-5187.

Pickett, C. L., Twiddy, E. M., Coker, C. and Holmes, R. K. (1989). Cloning, nucleotide sequence and hybridisation studies of the type IIb heat-labile enterotoxin gene of *Escherichia coli*. *J. Bacteriol.* **171**: 4945-4952.

Pieroni, P., Worobec, E. A., Paranchych, W. and Armstrong, G. D. (1988). Identification of a human erythrocyte receptor for colonization factor antigen I pili expressed by H10407 enterotoxigenic *Escherichia coli*. *Infect. Immun.* **56**: 1334-40.

Pohl, P., Lintermans, P., Van Muylen, K. and Schotte, M. (1982). Colibacilles enterotoxigenes du veau possedant in antigene d' attachment different de l'antigene K99. *Annu. Med. Vet.* **126**: 569-571.

Prasad, R., Chopra, A. K., Peterson, J. W., Pericas, R. and Houston, C. W. (1990). Biological and immunological characterisation of a cloned toxin-like enterotoxin from *Salmonella typhimurium*. *Microb. Pathog.* **9**: 315-329.

Rasheed, K. L., Guzman-Verduzco, L. and Kupersztoch, Y. M. (1990). Two precursors of the heat-stable enterotoxin of *Escherichia coli*: evidence of extracellular processing. *Mol. Microbiol.* **4**: 265-274.

Ratcliff, S. W., Luh, J., Ganesen, A. T., Behrens, B., Thompson, R., Montenegro, M., Morelli, G. and Trautner, T. A. (1979). The genome of *B. subtilis* phage SPP1; the arrangement of restriction endonuclease generated fragments. *Mol. Gen. Genet.* **168**: 165-172.

Rhen, M., Makela, P. H. and Korhonen, T. K. (1983). P-fimbriae of *Escherichia coli* are subject to phase variation. *FEMS Microbiol. Lett.* **19**: 267-271.

Rhen, M., Van Die, I., Rhen, V. and Bergmans, H. (1985). Comparison of the nucleotide sequences of the genes encoding the KS71A and $F7_1$ fimbrial antigens of uropathogenic *Escherichia coli*. *Eur. J. Biochem.* **151**: 573-577.

Rhen, M., Klemm, P. and Korhonen, T.K. (1986). Identification of two new haemagglutinins of *Escherichia coli*, N-acetyl-D-glucosamine-specific fimbriae and a blood group M-specific agglutinin, by cloning the corresponding genes in *Escherichia coli* K-12. *J. Bacteriol.* **168**: 1234-1242.

Riegman, N., Van Die, I., Leunissen, J., Hoekstra, W. and Bergmans, H. (1988). Biogenesis of F71 and F72 fimbriae of uropathogenic *Escherichia coli:* influence of the FsoF and FstFG proteins and localisation of the Fso/FstE protein. *Mol. Microbiol.* **2**: 73-80.

Roosendaal, B. (1987). A molecular study on the biogenesis of K99 fimbriae. *Thesis.*, Tessel Offset. Utrecht.

Roosendaal, B. and De Graaf, F. K. (1989). The nucleotide sequence of the fanD gene encoding the large outer membrane protein involved in the biosynthesis of K99 fimbriae. *Nucleic Acids Res.* **17**: 1263.

Roosendaal, B., Gaastra, W. and De Graaf, F. K. (1984). The nucleotide sequence of the gene encoding the K99 subunit of enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **22**: 253-258.

Roosendaal, B., Van Bergen en Henegouwen, M. P., Mooi, F. R., De Graaf, F. K. (1986). Regulatory aspects of the K99 fimbriae synthesis. Protein- carbohydrate interactions. *London Academic press* pp. 57-59.

Roosendaal, B., Jacobs, A. A. C., Rathman, P., Sondermayer, C., Stegehuis, F., Oudega, B. and De Graaf, F. K. (1987). Primary structure and subcellular

localisation of two fimbrial subunit-like proteins involved in the biosynthesis of K99 fibrillae. *Mol. Microbiol.* **1**: 211-213.

Roosendaal, B., Damoiseaux, J., Jordi, W. and De Graaf, F. K. (1989). Transcriptional organisation of the DNA region controlling expression of the K99 gene cluster. *Molec. Gen. Genet.* **215**: 250-256.

Rose, J. M., Houston, C. W., Coppenhaver, D. H., Dixon, J. D. and Kurosky, A. (1989). Purification and chemical characterisation of a cholera toxin-cross reactive cytolytic enterotoxin produced by a human isolate of *Aeromonas hydrophila*. *Infect. Immun.* **57**: 1165-1169.

Rosenberg, M. and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**: 319-353.

Rosenberg, M. and McKenney, C. K. (1983). Studying promoters and terminators by gene fusion. *Science* **222**: 734-739.

Ryoji, M., Hsia, K. and Kaji, A. (1983). Read-through translation. *Trends. Biochem. Sci.* **March**: 88-90.

Sakai, T., Sasakawa, C., Makino, S., Kamata, K. and Yoshikawa, M. (1986). Molecular cloning of a genetic determinant for congo-red binding ability which is essential for the virulence of *Shigella flexneri*. *Infect. Immun.* **51**: 476-482.

Sakai, T., Sasakawa, C., Makino, S. and Yoshikawa, M. (1987). DNA sequence and product analysis of the virF locus responsible for congo red binding and cell invasion in Shigella flexneri 2a. Infect. Immun., 54: 395-402.

Sakai, T., Sasakawa, C. and Yoshikawa, M. (1988). Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kDa virF protein. *Mol. Microbiol.* **2**: 589-597.

Salit, I. E., Vavougios, J. and Hofmann, T. (1983). Isolation and characterisation of *Escherichia coli* pili from diverse clinical sources. *Infect. Immun.* **42**: 755-762.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

Savelkoul, P. H. M, Willshaw, G. A., McConnell, M. M., Smith, H. R., Hamers, A. M., van-der-Zeijst, B. A. M. and Gaastra, W. (1990). Expression of CFA/I fimbriae is positively regulated. *Microb. Pathog.* **8**: 91-99.

Schmoll, T., Hoschutzky, H., Morschhauser, J., Lottspeich, F., Jann, K. and Hacker, J. (1989). Analysis of genes coding for the sialic acid-binding adhesin and two other minor fimbrial subunits of the S-fimbrial adhesin determinant of *Escherichia coli*. *Mol. Microbiol.* **3**: 1735-1744.

Schneider, K. and Beck, C. F. (1986). Promoter probe vectors for the analysis of divergently arranged promoters. *Gene* **42**: 37-48.

Schulz, S., Green, C. K., Yuen, P. S. T. and Garbers, D. L. (1990). Guanylase cyclase is a heat-stable enterotoxin receptor. *Cell* **63**: 941-948.

Scott, J. R., Wakefield, J. C., Russell, P. W., Orndorff, P. E. and Froehlich, B. J. (1992). CooB is required for assembly but not transport of CS1 pilin. *Mol. Microbiol.* **6**: 293-300.

Scotland, S. M., Gross, R. J. and Rowe, B. (1977). Serotype-related enterotoxigenicity in *Escherichia coli* O6.H16 and O148.H28. *J. Hyg. Camb.* **79**: 395-403.

Scotland, S. M., McConnell, M. M., Willshaw, G. A., Rowe, B. and Field, A. M. (1985). Properties of wild type strains of *Escherichia coli* which produce Colonization Factor Antigen II, and belong to serogroups other than O6. *J. Gen. Microbiol.* **131**: 2327-2333.

Seed, B. (1987). An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. *Nature (London)* **329**: 840-842.

Sherman, F. J., Formal, S. B. and Falkow, S. (1972). Plasmid associated enterotoxin production in a strain of *Escherichia coli* isolated from humans. *Infect. Immun.* **5**: 622-624.

Shinnick, T. M., Lerner, R. A. and Sutcliffe, J. G. (1981). Nucleotide sequence of moloney murine leukaemia virus. *Nature (London)* **293**: 543-548.

Silverman, M. and Simon, M. I. (1977). Bacterial flagella. *Annu. Rev. Microbiol.* **31**: 397-419.

Simons, B. L., Willemsen, P. T. J., Bakker, D., Roosendaal, B., De Graaf, F. K. and Oudega, B. (1990). Structure, localisation and function of FanF, a minor component of K99 fimbrillae of enterotoxigenic *Escherichia coli*. *Mol. Microbiol.* **4**: 2041-2050.

Simon, R., Priefer, U. and Puhler, A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in Gram negative bacteria. *Biotechnology* **1**: 784-791.

Sivaprasad, A. V., Jarvinen, R., Puspurs, A. and Egan, J. B. (1990). DNA replication studies with coliphage 186. III. A single phage gene is required for phage 186 replication. *J. Mol. Biol.* **213**: 449-463.

Sjoberg, P. O., Lindahl, M., Porath, J. and Wadstrom, T. (1988). Purification and characterisation of CS2, a sialic acid specific heamagglutinin of enterotoxigenic *Escherichia coli*. *Biochem. J.* **255**: 105-111.

Sjostrom, M., Wold, S., Wieslander, A. and Rilfors, L. (1987). Signal peptide aminoacid sequences in *Escherichia coli* contain information related to final protein localisation: a multivariate data analysis. *EMBO. J.* **6**: 823-831.

Smit, H., Gaastra, W., Kamerling, J. P., Vliegenthart, J. F. G. and De Graaf, F. K. (1984). Isolation and structural characterisation of the erythrocyte receptor for enterotoxigenic *Escherichia coli* K99 fimbrial adhesin. *Infect. Immun.* **46**: 578-584.

Smith, H. W. and Linggood, M. A. (1971). Observations on the pathogenic properties of the K88, Hly and Ent plasmids of *Escherichia coli* with particular reference to porcine diarrhoea. *J. Med. Microbiol.* **4**: 467-485.

Smith, H. W. and Linggood, M. A. (1972). Further observations on *Escherichia coli* enterotoxins with particular regard to those produced by atypical piglet strains and by calf and lamb strains: the transmissible nature of these enterotoxins and of a K antigen possessed by calf and lamb strains. *J. Med. Microbiol.* **5**: 243-250.

Smith, H. R., Willshaw, G. A. and Rowe, B. (1982). Mapping of a plasmid coding for colonization factor antigen I and heat stable enterotoxin production, isolated from an enterotoxigenic strain of *Escherichia coli*. *J. Bacteriol.* **149**: 264-275.

Smith, H. R., Scotland, S. M. and Rowe, B. (1983). Plasmids that encode for production of colonization factor antigen II and enterotoxin production in strains of *Escherichia coli*. *Infect. Immun.* **40**: 1236-1239.

Smyth, C. J. (1982). Two mannose-resistant haemagglutinins on enterotoxigenic *Escherichia coli* of serotype O6:K15:H16 or H- isolated from travellers' and infantile diarrhoea. *J. Gen. Microbiol.* **128**: 2081-2096.

Smyth, C. J. (1986). In Antigenic variation in *infectious diseases*. pp. 95-125. *In* Birkbeck, T.H. and Penn, C.W. (eds.) Special publications of the society for general Microbiology, No.19, IRL press, Oxford.

So, M. and McCarthy, B. J. (1980). Nucleotide sequence of the bacterial transposon Tn*1681* encoding a heat-labile (ST) toxin and its identification in enterotoxigenic *Escherichia coli* strains. *Proc. Natl. Acad. Sci. USA* **77**: 4011-4015.

So, M., Boyer, H. W., Betlach, M. and Falkow, S. (1976). Molecular cloning of an *Escherichia coli* plasmid determinant that encodes for the production of heat-stable enterotoxin. *J. Bacteriol.* **128**: 463-472.

So, M., Heffron, F. and McCarthy, B. J. (1979). The *E. coli* gene encoding heat-stable toxin is a bacterial transposon flanked by inverted repeats of IS*1*. *Nature (London)* **277**: 453-456.

Soderlind, O. and Mollby, R. (1979). Enterotoxins, O-group and K88 antigen in *Escherichia coli* from neonatal pigs with and without diarrhoea. *Infect. Immun.* **24**: 611-616.

Stormo, G. D., Schneider, T. D. and Gold, L. M. (1982). Characterization of translational initiation sites in *E. coli*. *Nucleic Acids Res.* **10**: 2971-2996.

Staden, R. (1980). A new method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* **8**: 3673-3694.

Staden, R. and McLachlan,A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Res.* **10**: 141-156.

Steven, A. C., Bisher, M. E., Trus, B. L., Thomas, D., Zhang, J. M, and Cowell, J. L. (1986). Helical structure of *Bordetella pertussis* fimbriae. *J. Bacteriol.* **167**: 968-974.

Stromberg, N., Marklund, B. I., Lund, B., Llver, D. and Hamers, A. (1990). Host-specificity of uropathogenic *Escherichia coli* depends on differences in binding specificity to Gal-α1-4Gal containing isoreceptors. *EMBO J.* **9**: 2001-2010.

Tabor, S. and Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**: 4767-4771.

Tacket, C. O., Maneval, D. R. and Levine, M. M. (1987). Purification, morphology and genetics of a new fimbrial putative colonization factor of enterotoxigenic *Escherichia coli* O159: H4. *Infect. Immun.* **55**: 1063-1069.

Taylor, W. E., Straus, D. B., Grossman, A. D., Burton, Z. F., Gross, C. A. and Burgess, R. R. (1984). Transcription from a heat-inducible promoter causes heat

shock regulation of the sigma subunit of *E. coli* RNA polymerase. *Cell* **38**: 371-381.

Teneberg, S., Willemsen, P., De Graaf, F. K. and Karlsson, K. A. (1990). Receptor-active glycolipids of epithelial cells of the small intestine of young and adult pigs in relation to infection with *Escherichia coli* K99. *FEBS Lett.* **263**: 10-14.

Tennent, J. M., Lindberg, F. and Normark, S. (1990a). Integrity of *Escherichia coli* P pili during biogenesis: Properties and role of PapJ. *Mol. Microbiol.*, **4**: 747-758

Tennent, J. M., Hultgren, S., Forsman, K., Göransson, M. and Marklund, B. I. (1990b). Genetics of adhesin expression in *Escherichia coli*. pp. 79-110. *In* B.H. Iglewski, V.C. Clark (eds.) Bacteria: Molecular basis of pathogenesis. Academic Press, San Diego.

Thiry, G., Clippe, A., Scarcez, T. and Petre, J. (1989). Cloning of DNA sequences encoding foreign peptides and their expression in the K88 pili. *Appl. Environ. Microbiol.* **55**: 984-993.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **71**: 4579

Thomas, L. V., Cravioto, A., Scotland, S. M. and Rowe, B. (1982). New fimbrial antigenic type (E8775) that may represent a colonization factor in enterotoxigenic *Escherichia coli* in humans. *Infect. Immun.* **35**: 1119-1124.

Thomas, L. V., McConnell, M. M., Rowe, B. and Field, A. M. (1985). The possession of three novel coli surface antigens by enterotoxigenic *Escherichia coli* strains positive for the putative colonization factor PCF8775. *J. Gen. Microbiol.* **131**: 2319-2326.

Tinoco, I. Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. and Gralla, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature (London)* **246**: 40-41.

Towbin, H., Stachelin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* 76: 4350-4354

Tsuji, T., Honda, T., Miwatani, T., Wakabayashi, S. and Matsubara, H. (1985). Analysis of receptor binding site in *Escherichia coli* enterotoxin. *J. Biol. Chem.* **260**: 8552-8558.

Twohig, J., Boylan, M. and Smyth, C. J. (1988). Expression of CS Fimbriae in *Escherichia coli* strains of Human and Animal Origin belonging to O-Serovar 6, K-serovar 15, or H-serovar 16. *FEMS Microbiol. Lett.* **56**: 327-330.

Uhlin, B. E., Norgren, M., Båga, M. and Normark, S. (1985a). Adhesion to human cells by *Escherichia coli* lacking the major subunit of digalactoside-specifc pilus-adhesin. *Proc. Natl. Acad. Sci. USA* **82**: 1800-1804.

Uhlin, B. E., Båga, M., Göransson, M., Lindberg, F. P., Lund, B., Norgren, M. and Normark, S. (1985b). Genes determining adhesin formation in uropathogenic *Escherichia coli*. pp. 163-178 In *Current Topics in Microbiology and Immunology* **118**: Springer, Berlin Heidelberg, New York.

Vaisanen-Rhen, V., Saarela, S. and Rhen, M. (1988). Mutations in cloned *Escherichia coli* P fimbriae genes that makes fimbriae production resistant to suppression by trimethoprim. *Microb. Pathog.* **4**: 369-377.

Van Embden, J. D. A., De Graaf, F. K., Schouls, L. M. and Teppema, J. S. (1980). Cloning and expression of a deoxyribonucleic acid fragment that encodes for the adhesive antigen K99. *Infect. Immun.* **29**: 1125-1133.

Van Die, I. and Bergmans, H. (1984). Nucleotide sequence of the gene encoding the F72 fimbrial subunit of a uropathogenic *Escherichia coli* strain. *Gene* **32**: 83-90.

Van Die, I., Van Megen, I., Hoekstra, W. and Bergmans, H. (1984a). Molecular organisation of the genes involved in the production of F72 fimbriae, causing a

mannose resistant haemagglutination of a uropathogenic *Escherichia coli*. *J. Bacteriol.* **167**: 407-410.

Van Die, I., Van Geffen., Hoekstra, W. and Bergmans, H. (1984b). TypeI C fimbriae of a uropathogenic *Escherichia coli* strain: cloning and characterisation of the genes involved in the expression of the 1C antigen and nucleotide sequence of the subunit gene. *Gene* **34**: 187-196.

Van Die, I., Spierings, G., Van Megen, I., Zuidweg, E., Hoekstra, W. and Bergmans, H. (1985). Cloning and genetic organisation of the gene cluster encoding F71 fimbriae of a uropathogenic *Escherichia coli* and comparison with the F72 gene cluster. *FEMS Microbiol. Lett.* **28**: 329-334.

Van Die, I., Zuidweg, E., Hoekstra, W. and Bergmans, H. (1986). The role of fimbriae of uropathogenic *Escherichia coli* as carriers of the adhesin involved in mannose resistant hemagglutination . *Microb. Pathog.* **1**: 51-56.

Van Die, I., Hoekstra, W. and Bergmans, H. (1987). Analysis of the primary structure of P fimbrillins of uropathogenic *Escherichia coli*. *Microb. Pathog.* **3**: 149-154.

Van Die, I., Wauben, M., Van Megen, I., Bergmans, H., Riegman, N., Hoekstra, W., Pouwels, P. and Enger-Valk, B. (1989). Genetic manipulation of major P-fimbrial subunits and its consequence for the formation of fimbriae. *J. Bacteriol.* **170**: 5870-5876.

Van Doorn, J., Oudega, B., Mooi, F. R. and De Graaf, F. K. (1982). Subcellular localisation of polypeptides involved in the biosynthesis of K88ab fimbriae. *FEMS Microbiol. Lett.* **13**: 99-104.

Van Verseveld, H. W., Bakker, P., Van der Woude, T., Terleth, C. and De Graaf, F. K. (1985). Production of fimbrial adhesins K99 and F41 by enterotoxigenic *Escherichia coli* as a function of growth-rate domain. *Infect. Immun.* **49**: 159-163.

Venkatesan, M. M., Buysse, J. M. and Kopecko, D. J. (1988). Characterization of invasion plasmid antigen genes (*ipaBCD*) from *Shigella flexneri*. *Proc. Natl. Acad. Sci. USA* **85**: 9317-9321.

Watts, T. H., Kay, C. M. and Paranchych, W. (1983). Spectral properties of three quaternary arrangements of *Pseudomonas* pilin. *Biochemistry* **22**: 3640-3646.

Weikel, C. S. and Guerrant, R. L. (1985). ST_B enterotoxin of *Escherichia coli*: cyclic nucleotide-independent secretion, pp. 94-115. *In* Microbial toxins and diarrhoeal disease. Evered R. and Whelan J. (ed.) Pittman, London.

Wenneras, C., Holmgren, J. and Svennerholm, A. M. (1990). The binding of colonization factor antigens of enterotoxigenic *Escherichia coli* to intestinal cell membrane proteins. *FEMS Microbiol. Lett.* **54**: 107-112.

Weiss, R. B. and Gallant, J. A. (1986). Frameshift suppression in aminoacyl tRNA-limited cells. *Genetics* **112**: 727-739.

Whipp, S. C. (1987). Protease degradation of *Escherichia coli* heat-stable, mouse-negative, pig-positive enterotoxin. *Infect. Immun.* **55**: 2057-2060.

White-Ziegler, C. A., Blyn, L. B., Braaten, B. A. and Low, D. A. (1990). Identification of an *Escherichia coli* genetic locus involved in thermoregulation of the pap operon. *J. Bacteriol.* **172**: 1775-1782.

Willshaw, G. A., Smith, H. R. and Rowe, B. (1983). Cloning of regions encoding colonization factor antigen 1 and heat-stable enterotoxin in *Escherichia coli*. *FEMS Microbiol. Lett.* **16**: 101-106.

Willshaw, G. A., Smith, H. R., McConnell, M. M. and Rowe, B. (1985). Expression of cloned plasmid regions encoding Colonization Factor Antigen I (CFA/I) in *Escherichia-coli*. *Plasmid* **13**: 8-16.

Willshaw, G. A., Smith, H. R., McConnell, M. M. and Rowe, B. (1988). Cloning of genes encoding Coli-Surface (CS) Antigens in Enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **49**: 473-478.

Willshaw, G. A., Smith, H. R., McConnell, M. M., Gaastra, W., Thomas, A., Hibberd, M. and Rowe, B. (1990a). Plasmid-encoded production of coli surface-associated antigen 1 (CS1) in a strain of *Escherichia coli* serotype O139.H28. *Microb. Pathog.* **9**: 1-11.

Willshaw, G. A., McConnell, M. M., Smith, H. R. and Rowe, B. (1990b). Structural and regulatory genes for coli surface associated antigen 4 (CS4) are encoded by separate plasmids in enterotoxigenic *Escherichia coli* strains of serotype 0.25.H42. *FEMS Microbiol. Lett.* **56**: 255-260.

Willshaw, G. A., Smith, H. R., McConnell, M. M. and Rowe, B. (1991). Cloning of regulator genes controlling fimbrial production by enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **82**: 125-130.

Wolf, M. K., Andrews, G. P., Tall, B. D., McConnell, M. M., Levine, M .M. and Boedeker, E. C. (1989). Characterisation of CS4 and CS6 antigenic components of PCF8775, a putative colonization factor complex from enterotoxigenic *Escherichia coli*. *Infect. Immun.* **57**: 164-173.

Yamamoto, T., Tamura, T. and Yokota, T. (1984a). Primary structure of heat-labile enterotoxin produced by *Escherichia coli* pathogenic for humans. *J. Biol. Chem.* **259**: 5037-5044.

Yamamoto, T., Nakazawa, T., Miyata, T., Kaji, A. and Yokota, T. (1984b). Evolution and structure of two ADP-ribosylation enterotoxins, *Escherichia coli* heat-labile toxin and *V.cholerae* toxin. *FEBS Lett.* **169**: 241-246.

Yanish-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119

Ziegler, V., Richards, K., Guilley, H., Jonard, G. and Putz, C. (1985). Cell-free translation of beet necrotic yellow vein virus: readthrough of the coat protein cistron. *J. Gen. Virol.* **66**: 2079-2087.