



Structural Studies of Afa/Dr Family of Fimbrial Proteins

Wei-Chao Lee

Division of Molecular Biosciences

Imperial College London

A thesis submitted for the degree of *Doctor of
Philosophy (PhD)*, and the *Diploma of Imperial College*

February 2012

Abstract

Pathogenic *Escherichia coli* and *Salmonella* spp. are Gram negative bacteria that inhabit the lower intestines of mammals and express virulence factors that are responsible for infecting millions of people with urinary tract infections, diarrhoea and food poisoning. These virulence factors, which target the host through many varied mechanisms, assist these organisms to overcome the host defences and colonize or invade the urinary or gastrointestinal tract.

This project focuses on Afa/Dr related fimbrial proteins from different strains of pathogenic *E. coli* and *Salmonella enteritidis* involved in the initial contact with host cells. A polymeric fibre of major subunits capped by a minor subunit make up the architecture of the fimbria and both are transported to the cell surface by the chaperone-usher system.

The structure of a minor subunit from Enteroaggregative *E. coli* (EAEC), AafB, was solved by X-ray crystallography at 3.0Å resolution. Potential ligands have been identified using glycoarrays and these interactions have been studied in more detail using nuclear magnetic resonance. Residues were identified that are important for interaction with heparin I-S. The structure of HdaB, another minor subunit from EAEC, was also solved by X-ray crystallography to 2.0 Å, although no ligand was identified. As expected these two homologous proteins show a very similar immunoglobulin-like fold consisting of 9 β -strands.

Structural work was also performed on SefD, a minor subunit from *S. enteritidis*. SefD crystals were not reproducible and molecular replacement failed to obtain phase information. It was possible to reproduce crystals by replacing 100% D₂O instead of H₂O into the crystallisation and sample buffer, although data collection and

processing showed a change in space group. Selenomethionine derivatives did not yield an anomalous signal and soaks using sodium iodide and 5-amino-2, 4, 6-triiodoisophthalic acid failed due to lack of incorporation.

Table of Contents

Abstract	2
Table of Contents	4
List of Figures	10
List of Tables	14
Acknowledgements	15
Dedication	16
Declaration of Originality	17
List of Abbreviations	18
Chapter 1: Biological Introduction	20
1.1 General Bacterial Overview	21
1.2 <i>Escherichia coli</i>	21
1.2.1 Pathogenic <i>E. coli</i>	22
1.2.1.1 Enteropathogenic <i>E. coli</i> (EPEC)	23
1.2.1.2 Enterotoxigenic <i>E. coli</i> (ETEC)	24
1.2.1.3 Enteroinvasive <i>E. coli</i> (EIEC)	24
1.2.1.4 Enterohemorrhagic <i>E. coli</i> (EHEC)	25
1.2.1.5 Enteroaggregative <i>E. coli</i> (EAEC)	25
1.2.1.6 Diffusely Adhering <i>E. coli</i> (DAEC)	27
1.3 Virulence Factors	27
1.4 Chaperone-Usher Pathway	29
1.4.1 Pilus Assembly	31
1.4.2 Chaperone-Subunit Complex	33
1.4.3 Donor Strands	35
1.5 Afa/Dr family	36
1.6 Afimbrial adhesin (Afa)	37
1.6.1 Genetic Organisation	38
1.6.2 Afa structures	39
1.6.3 Receptors	40
1.6.3.1 Type IV collagen and integrin $\alpha_5\beta_1$	41
1.6.3.2 Decay Accelerating Factor (DAF/CD55)	42
1.6.3.3 CEACAMs	42
1.6.4 Mechanism of infection	43
1.6.5 Roles of Afa adhesins in infection	44
1.7 Aggregative Adherence Fimbriae (Aaf)	46

1.7.1	Gene Cluster of Aaf-II	46
1.7.2	Receptors.....	48
1.7.3	Mechanism of infection	48
1.8	HUS-Associated Diffuse Adherent (Hda).....	48
1.9	<i>Salmonella</i>	49
1.9.1	Food Poisoning from <i>Salmonella</i>	50
1.10	<i>Salmonella enteritidis</i> Fimbriae (Sef)	50
1.10.1	Biogenesis of <i>sef</i> genes	50
1.10.2	Role of Infection	51
1.11	Project Aims.....	51
Chapter 2: Theories of Structural Techniques		53
2.1	Introduction	54
2.2	X-ray Crystallography.....	54
2.2.1	X-rays.....	55
2.2.1.1	Properties of X-rays.....	55
2.2.2	Crystals	56
2.2.2.1	Crystallization of Biological Molecules	56
2.2.2.2	Crystal Lattice and Symmetry	59
2.2.3	X-ray Diffraction by a Crystal	63
2.2.4	Data Collection and Processing	68
2.2.4.1	Data Collection	68
2.2.4.2	Auto-indexing and Integration of Reflections	69
2.2.4.3	Scaling Data.....	69
2.2.5	Phase Information	71
2.2.5.1	Patterson Function	71
2.2.5.2	Molecular Replacement.....	72
2.2.5.3	Anomalous Dispersion	75
2.2.5.4	Isomorphous Replacement	77
2.2.6	Model Refinement	78
2.2.6.1	Cross-validation.....	79
2.2.6.2	Least Squares and Maximum Likelihood Refinement	79
2.2.6.3	Specific Refinement Strategies.....	79
2.3	Nuclear Magnetic Resonance (NMR).....	83
2.3.1	Nuclear Spin Theory	83
2.3.2	Larmor Precession and Boltzmann Distribution.....	84
2.3.3	Pulse NMR.....	87
2.3.4	Coupling Effects	89
2.3.4.1	Scalar Coupling	89

2.3.4.2	Dipolar Coupling	90
2.3.5	Chemical Shift	90
2.3.5.1	Electron Shielding	91
2.3.5.2	Ring Current Shifts	91
2.3.6	Relaxation	92
2.3.6	One Dimensional NMR	93
2.3.7	Two Dimensional NMR.....	94
2.3.8	Three Dimensional NMR.....	95
2.4	Summary	98
Chapter Three: Materials and Methods.....		99
3.1	Recombinant DNA Techniques	100
3.1.1	Materials (DNA templates).....	100
3.1.2	Primers Design.....	101
3.1.3	Polymerase Chain Reaction (PCR).....	101
3.1.4	DNA Purification	102
3.1.4.1	PCR Purification.....	102
3.1.4.2	DNA Agarose Gels	102
3.1.4.3	DNA Extraction from Agarose Gels	103
3.1.5	Restriction Endonuclease Digestions.....	103
3.1.6	DNA Ligation	103
3.1.7	Transformation of <i>E. coli</i> Strains.....	104
3.1.8	Plasmid Preparation	105
3.1.9	Sequence Analysis	105
3.2	Protein Techniques	105
3.2.1	Protein Expression	105
3.2.2	Protein Purification	106
3.2.3	Protein Refolding	106
3.2.4	Protein Chromatography	106
3.2.5	SDS-PAGE	107
3.2.6	UV Spectroscopy	107
3.3	Nuclear Magnetic Resonance Experiments	108
3.3.1	Sample Buffer	108
3.3.2	NMR Data Acquisition	108
3.4	X-ray Crystallography	110
3.4.1	Pre-Crystallization Tests	110

3.4.2	Crystallization Screens.....	110
3.4.3	Cryoprotection	110
Chapter Four: Cloning, Expression, Purification and Crystallisation.....		112
4.1	Cloning.....	113
4.1.1	Designing of Constructs.....	113
4.1.2	Polymerase Chain Reaction	115
4.1.3	Ligation to PQE-30	115
4.2	Expression and Purification	116
4.2.1	Size Exclusion Chromatography.....	117
4.2.1.1	AfaD-VII	120
4.2.1.2	AafBA.....	120
4.2.1.3	SefA	121
4.2.1.4	HdaA.....	122
4.2.1.5	HdaB.....	122
4.2.1.6	HdaBA	122
4.3	One-Dimensional NMR	123
4.4	Crystallisation.....	124
4.4.1	AafB.....	125
4.4.2	HdaB	127
4.4.3	SefD	127
Chapter 5: Structural Determination and Analysis of AafB		130
5.1	Data Collection.....	131
5.2	Data Processing	131
5.3	Structural Determination	133
5.4	Structural Refinement and Modelling.....	135
5.5	Model Analysis and Validation.....	138
5.6	Structural Overview of AafB	141
5.6.1	Crystallography Structure of AafB	141
5.6.2	Comparisons of Molecules within the Asymmetric unit	142
5.6.3	Potential Binding Sites.....	144
5.7	Comparisons of AafB with Afa/Dr Homologues	146
5.7.1	Sequence Alignment	146
5.7.2	Secondary Structure Comparisons.....	146
5.7.3	Tertiary Structure Comparisons.....	148
5.8	Summary	150

Chapter 6 Functional studies of AafB.....	151
6.1 Interaction Screens	152
6.1.1 Glycosaminoglycan (GAG)	155
6.2 Nuclear Magnetic Resonance Studies	155
6.2.1 Suitability for NMR	155
6.2.2 2D NMR.....	157
6.2.3 NMR Backbone Assignments.....	158
6.2.3.1 HNCACB and CBCA(CO)NH.....	159
6.2.3.2 HN(CA)CO and HNCO.....	160
6.2.3.3 MARS Automatic Assignment.....	161
6.2.3.4 Multiple Species	162
6.2.4 Chemical Shift Index Analysis	163
6.2.5 Mapping Missing Regions on AafB Structure.....	164
6.3 Interaction Studies.....	166
6.3.1 ¹ H NMR of AafBA	166
6.3.2 Interaction of AafB with AafA	166
6.3.3 Titration of AafBA with Chondroitins.....	168
6.3.4 Titration of AafBA with Heparin.....	169
6.3.5 AafB Interacting with Heparin I-S.....	172
6.4 Summary	174
Chapter 7: Structural Determination and Analysis of HdaB	176
7.1 Data Collection.....	177
7.2 Data Processing.....	177
7.3 Structural Determination	179
7.4 Structural Refinement and Modelling	181
7.5 Model Analysis and Validation.....	185
7.6 Structural Overview of HdaB.....	188
7.6.1 Structure of HdaB	188
7.6.2 Comparisons of Molecules within the Asymmetric Unit	190
7.6.3 Potential Binding Sites.....	193
7.7 Comparisons of HdaB with Afa/Dr/Aaf Homologues	196
7.7.1 Sequence Alignment	196
7.7.2 Secondary Structure Comparisons.....	197
7.7.3 Tertiary Structure Comparisons.....	199
7.8 Summary	202

Chapter 8: Structural Studies of SefD.....	204
8.1 Data Collection.....	205
8.2 Data Processing.....	205
8.3 Alternative Crystallization Conditions.....	210
8.3.1 D ₂ O Crystallisation Conditions.....	210
8.3.2 Crystal Formation in D ₂ O	211
8.4 Crystals from D ₂ O Conditions	212
8.4.1 Data Collection	212
8.4.2 Data Processing.....	213
8.4.2.1 Native Dataset.....	213
8.4.2.2 Sodium Iodide and I3C Soaks	216
8.5 Selenomethionine Samples	217
8.6 Summary	218
8.6.1 Effect of D ₂ O	219
Chapter 9: Conclusions and Future Perspectives.....	221
9.1 Structural Studies of AafB and HdaB	222
9.1.1 Structure Determination of AafB and HdaB.....	222
9.1.2 Structural Comparisons of AafB and HdaB to Afa/Dr homologues....	223
9.2 Functional Studies of AafB and HdaB.....	223
9.2.1 Potential Binding Sites of AafB.....	223
9.2.2 NMR studies of AafB	224
9.2.3 Interaction Studies of AafB	224
9.2.4 Potential Binding site of HdaB	225
9.3 Effect of D ₂ O in SefD Structural Studies.....	225
9.4 Future Perspectives	226
9.4.1 Interaction Studies of AafB	226
9.4.2 Functional Studies of HdaB.....	226
9.4.3 Structural Studies of SefD	227
Appendices.....	228
Appendix 1: Sequence Alignment of AfaD-III homologues	228
Appendices 2: Primer sequences for building different constructs.....	229
References.....	230

List of Figures

Figure 1.1: Mechanisms of virulence by different diarrhoeagenic <i>E. coli</i>	29
Figure 1.2: Schematic representation of chaperone-usheer pathway.	31
Figure 1.3: Structure of FimD-FimC-FimH complex.....	32
Figure 1.4: Cartoon representation of Chaperone-Subunit complex of PapD-PapK...34	
Figure 1.5: Sequence Alignment of the donor strand from Pap subunits.	35
Figure 1.6: A schematic diagram for ‘zip-in, zip out’ mechanism for donor strand exchange (DSE).....	36
Figure 1.7: Genetic organisation of Afa/Dr operons.	38
Figure 1.8: Topology diagram of atomic structure of AfaE-III.	39
Figure 1.9: Images of AfaE-III filament.	40
Figure 1.10: Schematic representation of different receptors of Afa/Dr adhesin and invasin.....	41
Figure 1.11: The infection mechanism of DAEC.	43
Figure 1.12: Gene cluster of Aaf-II.....	47
Figure 1.13: Genetic organisation of the Hda operon.....	49
Figure 1.14: Operon of Sef and proposed mechanism of fimbria formation.....	51
Figure 1.15: Phylogram based on the alignment of AfaD-III homologues.	52
Figure 2.1: Electromagnetic Spectrum from radio waves to γ -rays.....	55
Figure 2.2: An illustration of the hanging-drop method of vapour diffusion.	58
Figure 2.3: Phase diagram illustrating protein solubility as a function of precipitant concentration.....	59
Figure 2.4: An example of unit cell from a crystal lattice.	60
Figure 2.5: Centring arrangement within a Unit cell.	61
Figure 2.6: Example of miller indices.....	63
Figure 2.7: Diagram representing Bragg’s law.....	64
Figure 2.8: Two-dimensional slice of the Ewald sphere.....	65
Figure 2.9: Definition of the Eulerian angles used to describe the rotation of a body.....	73
Figure 2.10: Argand diagram describing Friedel’s law.	76
Figure 2.11: Schematic representation of the magnetic dipole of a particle.....	84
Figure 2.12: Alignment of nuclear dipoles within a magnetic field.	84
Figure 2.13: The net magnetisation of nuclei in the magnetic field.	85
Figure 2.14: Difference in energy between the two spin states.	86

Figure 2.15: Effect of a pulse of RF on bulk magnetisation.....	87
Figure 2.16: A vector model representation of the pulse experiment.....	88
Figure 2.17: An example of J coupling effect between HA and HX in 2, 3-dimethylbutane.....	89
Figure 2.18: Longitudinal Relaxation Process.....	92
Figure 2.19: Transverse Relaxation Process.....	93
Figure 2.20: A schematic for 2D NMR Spectroscopy.....	94
Figure 2.21: The magnetisation transfer in the HNCACB experiment.	95
Figure 2.22: The magnetisation transfer in the CBCA (CO) NH experiment.	96
Figure 2.23: The magnetisation transfer in the HN (CA) CO experiment.....	97
Figure 2.24: The magnetisation transfer in the HNCO.....	97
Figure 3.1: The plasmid map pQE-30.....	104
Figure 4.1: Phylogram based on the alignment of AfaD-III homologues.	113
Figure 4.2: A schematic diagram demonstrating the design of constructs.	114
Figure 4.3: An image of 1.2% agrose gel confirming a succesful PCR of the constucts.	115
Figure 4.4: A SDS-PAGE showing samples from protein expression.	117
Figure 4.5: The results of Size exclusion chromatography.....	119
Figure 4.6: A SDS-PAGE of the samples observed in size exclusion chromatography.	120
Figure 4.7: SDS-PAGE of the SefA samples from size exclusion chromatography.	121
Figure 4.8: SDS-PAGE of HdaA, HdaB and HdaBA samples collected after size exclusion chromatography.....	123
Figure 4.9: One-dimensional spectra of purified protein samples.....	124
Figure 4.10: Images of crystals observed from a crystallization screen.	126
Figure 4.11: A typical drop with protein crystals from the crystallisation condition of ammonium citrate with PEG 3350.....	127
Figure 4.12: Images of crystals observed from different crystallization conditions and methods.....	129
Figure 5.1: The asymmetric unit of the AafB crystal..	134
Figure 5.2: Two different views of the crystal packing arrangement within the AafB crystal.....	134
Figure 5.3: An example of the electron density map quality of AafB.....	135
Figure 5.4: Ramachandran plots for the structure of AafB.....	139

Figure 5.5: Average B factor of AafB residues.	140
Figure 5.6: Cartoon representation and topology of the Structure of AafB.....	141
Figure 5.7: Superimposition of the two molecules in the asymmetric unit.	142
Figure 5.8: Adjacent contacts around the asymmetric unit.....	143
Figure 5.9: Crystals contacts between asymmetric units in the AafB unit cell.	144
Figure 5.10: The sulphate ion contact with AafB.	145
Figure 5.11: Sequence Alignment of AafB with Afa/Dr family.....	146
Figure 5.12: Comparisons of Secondary Structure of AafB with Afa/Dr Family.	147
Figure 5.13: Superimposition of AafB with AfaD and DraD.....	148
Figure 5.14: Comparisons of AafB with AfaD and DraD.	149
Figure 6.1: Microarray oligosaccharide screens of AafBA.	153
Figure 6.2: Microarray screens of Aaf with the GAG group of oligosaccharides.	154
Figure 6.3: 1D spectra of AafB before and after two weeks of incubation.	156
Figure 6.4: ^1H ^{15}N HSQC of AafB.....	157
Figure 6.5: Overlay of HSQC spectra of AafB at different temperatures.	159
Figure 6.6: Examples of the HNCACB and CBCA(CO)NH strips.	160
Figure 6.7: Examples of HN(CA)CO and HNCO strips.	161
Figure 6.8: Example of doubling of residues in spectra of AafB.	163
Figure 6.9: Secondary structure analyses using the Chemical Shift Index for AafB.	164
Figure 6.10: Missing residues from NMR assignment mapped onto the AafB crystal structure.....	165
Figure 6.11: Comparison of experimental secondary structures of AafB.	165
Figure 6.12: Overlaid HSQC spectra of AafBA with AafB and AafA.....	167
Figure 6.13: Residues of AafB affected by AafA.....	168
Figure 6.14: An example of titration of AafBA with chondroitin sulphate C (CSC).	169
Figure 6.15: Structures of Heparin derivatives.....	172
Figure 6.16: Titration of AafBA with Heparin III-S and Heparin I-A at 1:40 ratio..	170
Figure 6.17: Titration of AafBA with Heparin I-S.	171
Figure 6.18: Titration of AafB with Heparin I-S.	173
Figure 7.1: The asymmetric unit of the HdaB crystal.....	180
Figure 7.2: Two different views of the crystal packing arrangement within the HdaB crystal.....	180

Figure 7.3: An example of the electron density map quality of HdaB.	181
Figure 7.4: Results from TLSMD.	184
Figure 7.5: Ramachandran plots for the structure of HdaB.	186
Figure 7.6: Average B factor of HdaB residues.	187
Figure 7.7: Ribbon representation and topology of the Structure of HdaB.	189
Figure 7.8: Superimposition of the two molecules in the asymmetric unit.	190
Figure 7.9: Adjacent contacts around the asymmetric unit.	191
Figure 7.10: Crystals contacts between asymmetric units in the HdaB unit cell.	192
Figure 7.11: Alignment of AafB with HdaB.	194
Figure 7.12: The contact regions between citrate ions and HdaB.	195
Figure 7.13: Sequence Alignment of HdaB with Afa/Dr/Aaf.	196
Figure 7.14: Comparisons of the secondary structure of HdaB with Afa/Dr/AafB... ..	198
Figure 7.15: Superimposition of HdaB with AfaD, DraD and AafB.	199
Figure 7.16: Superimposeition of HdaB with DraD and AafB.	201
Figure 7.17: Superimposeition of HdaB with AfaD.	202
Figure 8.1: Systematic absences along the <i>00l</i> reflections for C222.	207
Figure 8.2: Self-rotation function of SefD crystals.	208
Figure 8.3: Images of crystals grown in different concentrations of D ₂ O.	212
Figure 8.4: Patterson Self-Rotation map of SefD indexed in spacegroup P222.	214
Figure 8.5: Native Patterson map of P222.	214
Figure 8.6: Asymmetric units in unit cell of C222 and P222.	220

List of Tables

Table 1.1: Different strains of intestinal <i>E. coli</i> and their associated pathologies and hosts.	23
Table 1.2: Characteristics of Afa/Dr family.	37
Table 2.1: Definition the seven crystal systems.....	62
Table 3.1: PCR Reaction recipes	102
Table 3.2: Digestion buffer used in double digestion with different restriction enzymes.....	103
Table 3.3: NMR experiments and acquisition parameters.....	109
Table 3.4: Composition of different cryoprotection solutions.....	111
Table 4.1: Range of optimisation conditions for crystallisation of AafB	126
Table 4.2: A range of optimisation conditions for crystallisation of HdaB.....	127
Table 4.3: A range of optimisation conditions for crystallisation of SefD.....	128
Table 5.1: Summary of data collection and scaling statistics for AafB.....	132
Table 5.2: Summary of Matthews Coefficient of AafB.....	132
Table 5.3: Z-score of the two molecules found from PHASER.	133
Table 5.4: Refinement statistics for the final refined AafB model.	137
Table 7.1: Summary of data collection and scaling statistics for HdaB.	178
Table 7.2 Summary of Matthews Coefficient of HdaB.	179
Table 7.3: Z-score of the two molecules found from PHASER.	179
Table 7.4: Refinement statistics for the final refined HdaB model.	183
Table 8.1: Summary of data collection and scaling statistics for SefD.	206
Table 8.2: Summary of Matthews Coefficient for SefD crystal.	207
Table 8.3: Z-score of the partial solutions found from PHASER.....	209
Table 8.4: Conditions for D ₂ O crystallisation screens.....	211
Table 8.5: Summary of data collection and scaling statics for SefD.	215
Table 8.6: Matthews Coefficient of SefD in D ₂ O.....	216
Table 8.7: Cell dimensions and angles of P222 and C222.	219

Acknowledgements

I would like to thank my supervisor Prof. Steve Matthews for giving me the opportunity to do this PhD. He has been very supportive for my research and always pointed me the right direction when I am lost in my research or facing difficulties in life. In particular, he has offered me a technician position which provided financial stability during my PhD.

Dr. Ernesto Cota helped me so much both in the lab and outside the lab. He has been very patient and spent so much time teaching me how to conduct experiments. I think it is fair to say almost all the techniques I have learnt in the wet lab came from him. I also like to thank him for his guidance which is like a beacon showing me the way out of the dilemma in life.

Big thanks to Dr. James Garnett, whom introduced me to the world of X-ray crystallography. He has been patiently explaining and teaching me many crystallography concepts and experiments even I have been asking the same or silly questions over and over again. He made all the synchrotron trips memorable and fun and without his help this PhD would not go as smoothly.

I would like to thank Dr. Pete Simpson for his support in helping me with running NMR experiments. He has made many of the lab events fun and interesting. He is a very busy person but he is always there to help, no matter how small the problem is.

I also want to thank all present and former members in the Matthew/Cota groups. Jon, Maud, Yingqi, Esther and Paula have made my experience in the lab very pleasant and enjoyable. Jan has been a great help especially in his wickedness with Linux application which made life so much easier. Furthermore, Savvas, Kov, Yang, Bing, Rob, Rex, Minhao, Ben, Llyod and Seb whom have made my non-scientific life much more fun, excited and memorable. I will never forget all the fun we have, particularly the exhausted but irreplaceable climbing experiences in CCPN, the sweating basketball games, delicious meals and interesting games we have played.

Finally, I would like to thank my family especially my mother, cousins and uncles who have been very understanding and supportive for me, no matter what I have done.

Dedication

Dedicating this thesis to my mother

Declaration of Originality

I hereby declare that the results presented in this thesis are the generated by my own work except where specifically indicated in the text. The information derived from others has been listed in references.

Signed

List of Abbreviations

	Mean B-factor
A/E	Attaching and Effacing
Aaf	Aggregative adherence fimbriae
Afa	Afimbrial adhesin
ATP	Adenosine triphosphate
BFP	Bundle-forming pilus
BLAST	Basic local alignment search tool
bp	Base pairs
CCP	Complement control protein
CEA	Carcinoembryonic antigen
CEACAM	Carcinoembryonic antigen cell adhesion molecules
CFs	Colonisation factors
CLA	Chain-like adherence
CSI	Chemical shift index
CTD	C-terminal domain
C-terminal	Carboxy-terminal
DAEC	Diffusely adherent <i>E. coli</i>
DAF	Decay accelerating factor
DNA	deoxyribonucleic Acid
DSC	Donor strand complementation
DSE	Donor strand exchange
<i>E. coli</i>	<i>Escherichia coli</i>
EAEC	Enteragggregative <i>E. coli</i>
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohaemorrhagic <i>E. coli</i>
EIEC	Enteroinvasive <i>E. coli</i>
EM	Electron microscopic
EPEC	Enteropathogenic <i>E. coli</i>

ETEC	Enterotoxigenic <i>E. coli</i>
FID	Free induction decay
FT	Fourier transformation
Hda	HUS-associated diffuse adherent
HSQC	Heteronuclear single-quantum coherence
HUS	Haemolytic uremic syndrome
Hz	Hertz
Ig-like	Immunoglobuline-like
IPTG	Isopropyl- β -D-1-thiogalactopyranoside
kDa	Kilodalton
LEE	Locus of enterocyte effacement
LTs	Heat-labile enterotoxins
MNEC	Meningitis-associated <i>E. coli</i>
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
PAI	Pathogenicity island
PCR	Polymerase chain reaction
PDB	Protein data bank
Ppm	Parts per million
RF	Radio frequency
RMSD	Root mean square deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sef	<i>Salmonella enteritidis</i> fimbriae
STs	Heat-stable enterotoxins
Stx	Shiga toxin
UPEC	Uropathogenic <i>E. coli</i>
UTI	Urinary tract infections

Chapter 1: Biological Introduction

1.1 General Bacterial Overview

Bacteria are a large group of prokaryotic microorganisms. They replicate rapidly through dividing in two by binary fission and under optimal conditions a single cell can divide every 20 minutes. This allows them to inhabit a wide range of environments including: soil, water, ocean depths, salt brines, hot acid springs, organic matter and the living bodies of animals and plants.

They display a wide range of shapes such as spherical (cocci), rod (bacilli)-shapes and filamentous, and are usually 0.5-5.0 μm in length (Robertson *et al.* 1975). The ability of the organism to uptake nutrients from the environment, attach to surfaces and swim depends greatly on the morphology of the bacteria which results from the structure of the cellular peptidoglycan cell wall and the intracellular cytoskeleton (Fritz *et al.* 2004; Cabeen and Jacobs-Wagner 2005; Young 2006; Wanger *et al.* 2008).

On the surface of the cell, flagella (~ 20 nm in diameter) can be found which are used for motility (Kojima and Blair 2004). In addition, fimbriae which are fine filament proteins (~ 2 -10 nm in diameter) are distributed over the surface of the cell and are involved in attachment to solid surfaces and in some cases are essential for the virulence for bacterial pathogens (Beachey 1981; Silverman 1997).

1.2 *Escherichia coli*

E. coli is a Gram-negative bacterium that inhabits the lower intestines of mammals. In the majority of instances *E. coli* are beneficial to humans (e.g. in waste processing, vitamin K production, food absorption in the large intestine and preventing the establishment of pathogenic bacteria within the intestinal tract). As they are able to survive outside the body, they are commonly used as an indicator of fecal

contamination and to monitor water quality. *E. coli* are also a model system for many biological studies as they can be grown readily and have relatively simple genetics which can be easily manipulated.

Some *E. coli* strains are pathogenic and contain virulence factors that are responsible for a range of different diseases in humans. These virulence factors, which target the host through many varied mechanisms, assist the organism to overcome the host defences and colonize or invade the urinary or gastrointestinal tract (Johnson 1991; Natuzzi *et al.* 1993; D'Orazio and Collins 1998; Nataro and Kaper 1998; Kaper *et al.* 2004).

1.2.1 Pathogenic *E. coli*

The virulence of pathogenic *E. coli* is frequently encoded on genetic elements that can be mobilized into different strains to create novel combinations of virulence factors. However, some genetic elements may have evolved within the genome of the pathogen. As with the survival of the fittest theory, only the most successful combinations of virulence factors have persisted to become specific and pass on to further generations. In present day there are three clinical identified pathotypes: enteric/diarrhoeal disease, urinary tract infections (UTIs) and sepsis/meningitis (Kaper *et al.* 2004).

E. coli pathogens can be further divided into at least six groups based on their serotypes and the mechanism by which the disease is thought to be induced: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) (Table 1.1) (Kaper *et al.* 2004). UTIs are the most common extraintestinal *E. coli* infections and are caused by

uropathogenic *E. coli* (UPEC). Meningitis-associated *E. coli* (MNEC) is responsible for meningitis and sepsis and is an increasingly common cause of extraintestinal infections.

<i>E. coli</i> strains	Symptoms	Host
Enteropathogenic <i>E. coli</i> (EPEC)	Acute or persistent watery diarrhoea	Human, pig, rabbit, calf, dog
Enterotoxigenic <i>E. coli</i> (ETEC)	Watery diarrhoea	Human, pig, beef, sheep
Enteroinvasive <i>E. coli</i> (EIEC)	Watery diarrhoea	Human
Enterohemorrhagic <i>E. coli</i> (EHEC)	Acute diarrhoea, hemorrhagic colitis	Human, calf
Enteraggregative <i>E. coli</i> (EAEC)	Watery persistent diarrhoea	Human
Diffusely Adhering <i>E. coli</i> (DAEC)	Acute or persistent watery diarrhoea	Human, calf

Table 1.1: Different strains of intestinal *E. coli* and their associated pathologies and hosts.

1.2.1.1 Enteropathogenic *E. coli* (EPEC)

EPEC was first isolated from a group of children in the UK in 1945 during a large outbreak of diarrhoea. Although large outbreaks of infant diarrhoea due to EPEC have almost disappeared from industrialized countries, EPEC still remains an important cause of potentially fatal infant diarrhoea in developing countries (Nataro and Kaper 1998). In the past, the mechanisms by which EPEC caused diarrhoea were unknown and could only be identified by serotyping. Since 1979, many studies have

dramatically advanced our understanding of EPEC diarrhoea and EPEC has now become one of the best understood pathogenic *E. coli*. EPEC have the characteristic known as attaching and effacing (A/E), in which the bacteria attach to intestinal epithelial cells and cause accumulation of polymerized actin beneath the adherent bacteria (McDaniel *et al.* 1995). This ability is encoded by the pathogenicity island (PAI) known as the locus of enterocyte effacement (LEE) and can be found in other human and animal pathogens including EHEC and rabbit EPEC (McDaniel *et al.* 1995).

1.2.1.2 Enterotoxigenic *E. coli* (ETEC)

ETEC causes watery diarrhoea which can range in severity from a mild, self-limiting disease to a severe purging one. ETEC is responsible for many cases of childhood diarrhoea in the developing world and is the main cause of travellers' diarrhoea (Nataro *et al.* 1998). ETEC colonizes the surface of the small bowel mucosa and produces enterotoxins, which give rise to intestinal secretion. Colonization is mediated by fimbrial colonization factors (CFs) which include more than 20 antigens (Wolf 1997). ETEC can express heat labile enterotoxins (LTs), heat-stable enterotoxins (STs) or both (Spangler 1992). ETEC is largely a pathogen of developing countries especially in those with a low rate of colon cancer. Pitari *et al.* (Pitari *et al.* 2003) have reported that STa, which is secreted by ETEC, can suppress colon cancer cell proliferation through a guanylyl cyclase C-mediated signalling cascade.

1.2.1.3 Enteroinvasive *E. coli* (EIEC)

EIEC are biochemically, genetically and pathogenically closely related to *Shigella* spp. and unlike other enterobacteria, they lack functional curli genes (Pupo *et al.* 2000; Wei *et al.* 2003), but clinical differences can be used to distinguish them. EIEC can

cause an invasive inflammatory colitis and dysentery, but in most cases EIEC is responsible for watery diarrhoea which is indistinguishable from infection by other pathogenic *E. coli* (Nataro *et al.* 1998). EIEC can only be distinguished from *Shigella* by a few biochemical tests but the two organisms share many essential virulence factors such as IpaA, IpaB, IpaC and IpgD which belong to the plasmid-borne type III secretion system (Sansonetti *et al.* 2000; Tran Van Nhieu *et al.* 2000). EIEC infection is thought to show an inflammatory colitis, although many patients are diagnosed with small bowel syndrome (Kaper *et al.* 2004).

1.2.1.4 Enterohemorrhagic *E. coli* (EHEC)

EHEC was first identified in 1982 and responsible for bloody diarrhoea (haemorrhagic colitis), non-bloody diarrhoea and haemolytic uremic syndrome (HUS). The initial outbreaks were associated with consumption of undercooked hamburgers and subsequently, a wide variety of foods have been associated with the disease, including sausages, unpasteurized milk, lettuce, cantaloupe melon, apple juice and radish sprouts (Wells *et al.* 1983; Brooks *et al.* 2005; Caprioli *et al.* 2005). In addition, EHEC needs an extremely low bacterial count for infection to take hold (estimated to be <100 cells) (Kaper *et al.* 2004) and therefore further outbreaks have occurred which have originated from drinking water, person-to-person transmission and zoo and farm visitations (Rowe *et al.* 1993; Rashid *et al.* 2006). A recent report suggested that a potential airborne transmission in a closed space such as a building is possible (Varma *et al.* 2003).

1.2.1.5 Enteroaggregative *E. coli* (EAEC)

EAEC strains are characteristic by their aggregative adherence or as also known as the stacked-brick (Nataro *et al.* 1987; Nataro and Kaper 1998; Nataro *et al.* 1998) and

chain-like adherence (CLA) (Gioppo *et al.* 2000). They are associated with persistent diarrhoea in children (Mathewson *et al.* 1987; Nataro *et al.* 1987; Bhan *et al.* 1989; Cravioto *et al.* 1991), acute diarrhoea (Scaletsky *et al.* 2002) and act as a causative agent of diarrhoeal disease in human immunodeficiency virus (HIV)-infected adults in the developed world (Wanke *et al.* 1998; Wanke *et al.* 1998; Durrer *et al.* 2000).

Epidemiological studies have shown that in developing countries the source of EAEC infection associated with food and water (Adachi *et al.* 2002; Falcao *et al.* 2004). In Mexico, 51% of children less than 2 years old and in Brazil, 20% of children less than 5 years old, with persistent diarrhoea, were associated with EAEC (Cravioto *et al.* 1991; Wanke *et al.* 1991). Furthermore, a high frequency of severe acute diarrhoea caused by EAEC was found in India and Venezuela (Paul *et al.* 1994; Kang *et al.* 1995; Gonzalez *et al.* 1997).

In developed countries such as England and Germany, EAEC were often found in stool samples from children with persistent and acute diarrhoea (Chan *et al.* 1994; Huppertz *et al.* 1997). With many people travelling from developed to developing countries, traveller's diarrhoea has become a common disorder and Adachi *et al.* have shown EAEC prevalence in 26% of traveller's diarrhoea in three different regions of the world (Adachi *et al.* 2001).

In the last few years, EAEC has become a potential cause for diarrhoea in HIV-positive patients as EAEC strains has been detected from diarrhoeal stools from these patients (Mayer and Wanke 1995; Polotsky *et al.* 1997; Wanke *et al.* 1998; Mossoro *et al.* 2002). In addition, the recent outbreak across Europe was caused by a hybrid strain of EAEC and EHEC (Bielaszewska *et al.* 2011; Frank *et al.* 2011; Scheutz *et al.* 2011; Struelens *et al.* 2011)

1.2.1.6 Diffusely Adhering *E. coli* (DAEC)

DAEC stains have been identified as a group of pathogens based on their diffuse pattern of adherence on cultured epithelial Hep-2 and HeLa cells (Scaletsky *et al.* 1984; Nataro *et al.* 1985; Nataro and Kaper 1998) and can be further divided into two classes: typical and atypical strains.

Studies have shown that all Afa/Dr DAEC adhesins act as virulence factors and are only present in pathogenic strains of *E. coli* and their sequences are homologous to putative virulence genes (Bilge *et al.* 1989; Blanc-Potard *et al.* 2002). The typical strains of DAEC are capable of harbouring Afa/Dr adhesins and are responsible for causing urinary tract infections (UTIs) and other various enteric infections (Archambaud *et al.* 1988; Daigle *et al.* 1994; D'Orazio and Collins 1998; Nowicki *et al.* 2001).

The second class (atypical) of DAEC causes infantile diarrhoea and express an adhesion involved in diffuse adherence (AIDA-I) (Benz and Schmidt 1989; Benz and Schmidt 1992; Benz and Schmidt 1992; Benz and Schmidt 1993). These strains also contain one or more homologues of the locus of the enterocyte effacement characteristic of EPEC, which contribute to the pathogenic potential of these strains. The diarrhoeagenic *E. coli* strains have been shown to secrete proteins which regulated by environmental parameters such as growth medium, temperature, pH and ion concentration (Beinke *et al.* 1998).

1.3 Virulence Factors

Pathogenic enteric bacteria strains produce specific virulence factors which allow them to recognise and colonize host cells such as epithelial cells, specific intestinal cells and others involved in immune responses, such as macrophages (Baumler *et al.*

1996; Baumler *et al.* 1996; Edwards *et al.* 2000). These adhesins form distinct morphological structures known as fimbriae or pili. Some pili are also known as type IV pili and are generally 6-7 nm in diameter. Pili adhere to a solid surface or other bacteria and promote twitching motility. Other pili (fimbriae) are shorter with rod-like structures 5–10 nm in diameter and are required to initiate colonisation and biofilm formation.

The Afa/Dr family of adhesins are produced by many diarrhoeagenic and uropathogenic *E. coli* as afimbrial adhesins, although they are fine fimbrial structures which are difficult to visualize (Keller *et al.* 2002). For example, the Afa/Dr family of adhesins from DAEC can recognise Decay Accelerating Factor (DAF, also known as CD55), which results in activation of phosphatidylinositol 3-kinase (PI-3-kinase) and cell-surface expression of the major histocompatibility complex (MHC) class I-related molecule MICA (Tieng *et al.* 2002) (Fig. 1.1). Another Afa/Dr family, Aaf is produced by EAEC and is responsible for the initial adhesion to host cells and the first stage of biofilm formation, which facilitates the secretion of cytotoxins and enterotoxins such as ShET1, Pic, EAST1 and Pet (Nataro *et al.* 1998) (Fig. 1.1) .

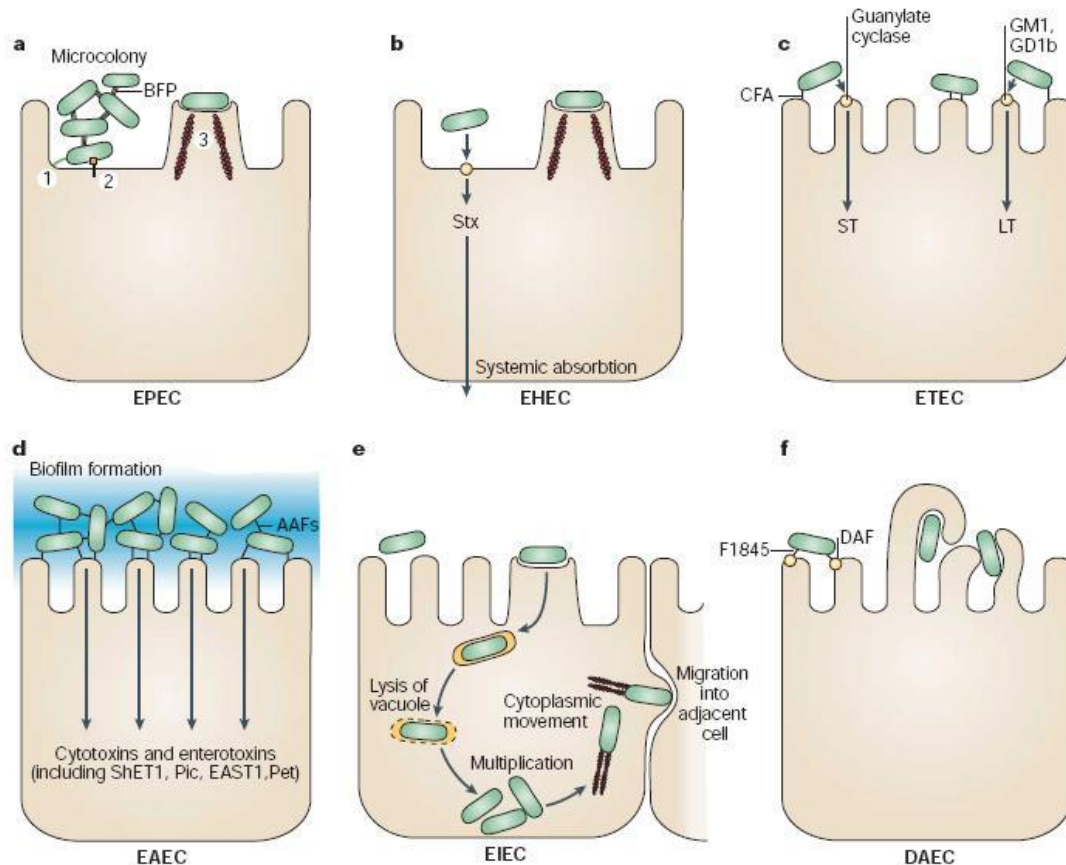


Figure 1.1: Mechanisms of virulence by different diarrhoeagenic *E. coli*. (a) EPEC adheres to the host cells which destroys the normal microvillar and causes rearrangement of the cytoskeleton and leads to inflammation and diarrhoea in the small bowel. (b) EHEC and (c) ETEC adhere to host cells in the colon and small bowel, respectively and secrete enterotoxins. (d) Using Aaf proteins EAEC adheres to the host cell and forms a biofilm which facilitates the secretion of cytotoxins and enterotoxins. (e) It has been suggested that EIEC enters the cell laterally through the epithelium cell by either a direct cell-to-cell spread or through exiting and re-entering the basolateral plasma membrane. (f) DAEC adhere to host cells by producing extracellular proteins such as Afa/Dr adhesins which recognise receptors (i.e. DAF) on the surface of the host cell and further trigger a cascade of invasion events. This image is adapted (Kaper *et al.* 2004).

1.4 Chaperone-Usher Pathway

Gram-negative bacteria, including many pathogens (i.e. *E. coli* and *Salmonella*) have conserved pathways to produce pili or fimbriae such as the chaperone -usher pathway. This pathway enables the bacteria to generate diverse multi-subunit fibres on the cell surface with different morphologies ranging from thick pilus rods to very thin fimbriae. These fibres play an important role in adhesion to the host cell, facilitate invasion, confer antibiotic resistance and the ability to form biofilms. Typically, the

chaperone-usher pathway is encoded in a gene cluster which includes periplasmic chaperones, an outer-membrane usher protein, a major fimbrial subunit (adhesin) and a minor fimbrial subunit (invasin) (Servin 2005).

Initially, the fimbrial subunits such as AfaE and AfaD from DAEC enter the periplasm via the Sec translocation machinery which is driven by ATP (Manting and Driessen 2000) (Fig.1.2). Subsequently, the subunits interact with the chaperone to form a chaperone-subunit complex, but in the absence of chaperones, the subunits misfold, aggregate and are proteolytically degraded by the protease DegP (Hultgren *et al.* 1989; Slonim *et al.* 1992; Jones *et al.* 1997; Barnhart *et al.* 2000; Vetsch *et al.* 2004). Therefore, formation of this complex allows the subunit to be efficiently folded and stabilised which prevents degradation or the formation of pre-mature fibres in the periplasm (Kuehn *et al.* 1991; Kuehn *et al.* 1993; Bullitt *et al.* 1996; Jones *et al.* 1997; Soto *et al.* 1998; Barnhart *et al.* 2000; Sauer *et al.* 2002). The chaperone-subunit complexes are targeted to the outer-membrane usher (Dodson *et al.* 1993) which facilitates the opening up of the chaperone-subunit complex hence exposing the interactive surface for polymerisation of the subunit to occur. The outer-membrane usher is a dimer and forms two ring shapes with a pore in the centre which allows the subunits to pass through and form polymers (Li *et al.* 2004).

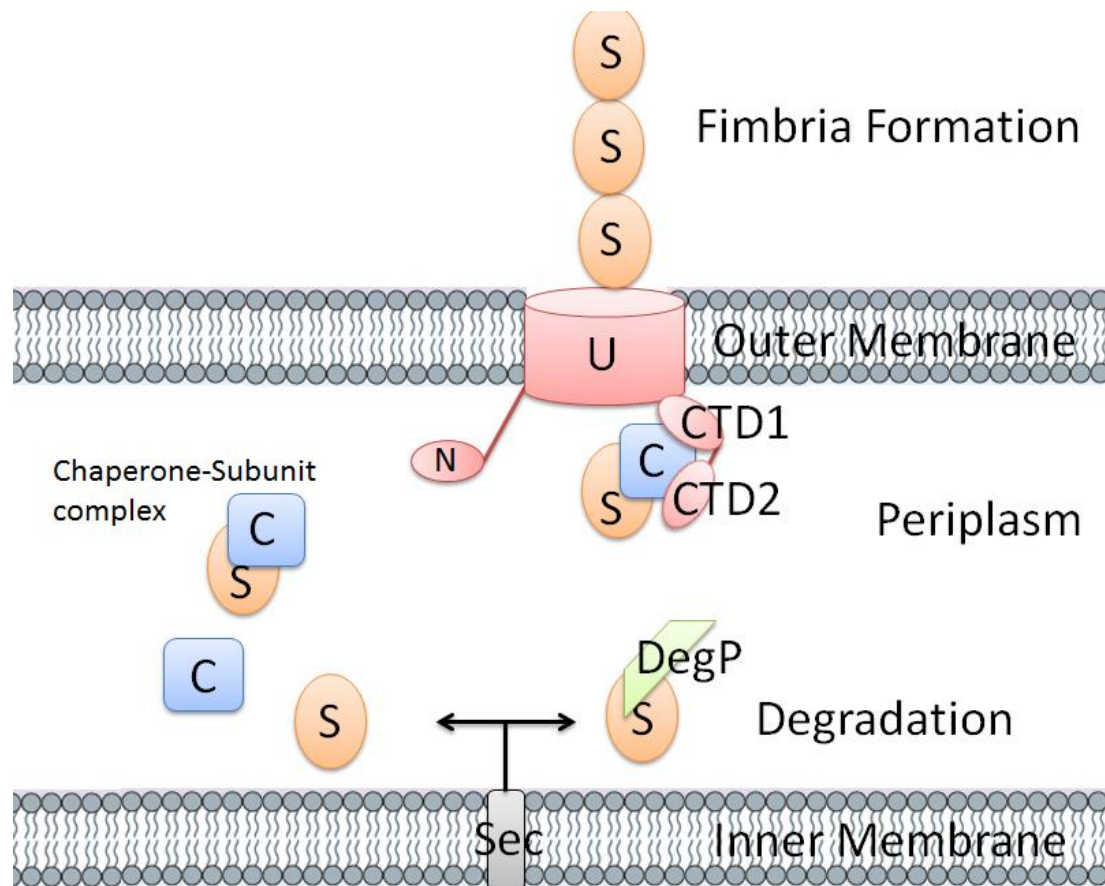


Figure 1.2: Schematic representation of chaperone-usher pathway. Subunits (S) are secreted through the Sec system to the periplasm and the chaperone (C) is needed to form a stable chaperone-subunit complex. In the absence of chaperone the subunits will misfold and led to degradation by DegP. The chaperone-subunit complex interacts with the N-terminal (N) of the usher protein (U) then passed to the C-terminal (CTD1 and CTD2) where the disassociation of the chaperone-subunit complex and the formation of fimbrial subunit-subunit complex take place for transportation to the surface (Sauer *et al.* 2000).

1.4.1 Pilus Assembly

The usher consists of around 800 residues including 4 domains, an N-terminal (~125 residue periplasmic) domain, a C-terminal (~170 residue periplasmic domain), large central domain (~500 residues) and a conserved plug domain (~110 residues) (Thanassi *et al.* 2002; Nishiyama *et al.* 2003; Capitani *et al.* 2006). The N-terminal domain binds to the chaperone-subunit complexes and most interfaces are in the chaperone but a small part of the interface is involved in subunit-specific interaction (Dodson *et al.* 1993; Saulino *et al.* 1998; Nishiyama *et al.* 2003). The C-terminal region consists of two domains (CTD1 and CTD2) which interact with the chaperone-

subunit complex. At this position, the usher-chaperone-subunit complex is allowed to sit and wait for the arrival of the new chaperone-subunit complex which is delivered by the N-terminal domain on the usher (Fig. 1.3) (So and Thanassi 2006; Huang *et al.* 2009; Phan *et al.* 2011). The N-terminal domain of the usher protein specifically picks up the appropriate chaperone-subunit complex and the CTDs provide a venue for the formation of subunit-subunit complex leading to the elongation of the subunits and hence fimbrial formation.

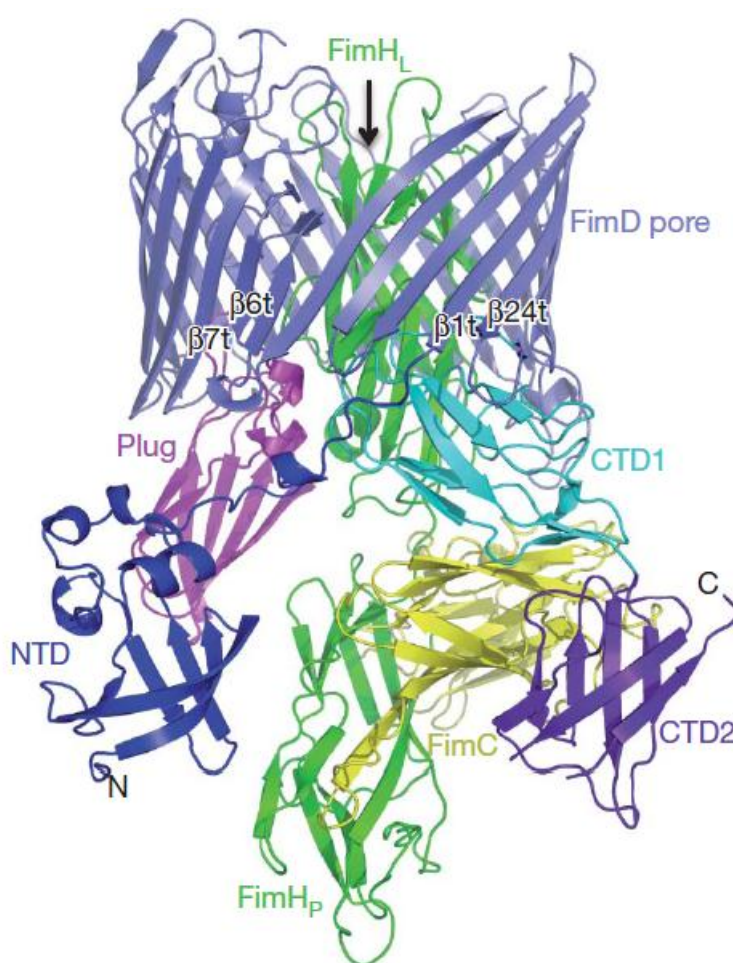


Figure 1.3: Structure of FimD-FimC-FimH complex. Cartoon representation of type I pili from uropathogenic *E. coli*. FimD-FimC-FimH complex with FimH (subunit) in green, FimC (chaperone) in yellow and FimD (usher) N-terminal domain (NTD), β -barrel, plug, C-terminal domain 1 and 2 (CTD1 and CTD2) in blue, slate, magenta, cyan and purple, respectively. The labelled $\beta 1t$, $\beta 6t$, $\beta 7t$ and $\beta 24t$ are the β -strands connecting the usher barrel to NTD, plug and CTDs, respectively. The image is adapted from Phan *et al* 2011 (Phan *et al.* 2011).

1.4.2 Chaperone-Subunit Complex

The chaperones and the subunits involved in fimbrial formation have an immunoglobulin-like fold (Ig-like). Subunits assembled by the chaperone-usher pathway have a conserved structural feature which was first determined by the crystal structure of the FimC-FimH and PapD-PapK chaperone-subunit complexes (Choudhury *et al.* 1999; Pellecchia *et al.* 1999; Sauer *et al.* 1999). In these complexes, a motif of four conserved alternating hydrophobic residues form a β -strand (G1 strand in the Pap system) and insert parallel into a groove in the hydrophobic core of the subunit via donor strand complementation (Choudhury *et al.* 1999; Sauer *et al.* 1999). The chaperones have conserved residues, Arg 8 and Lys 112 (numbering for Pap) which are essential for the formation of chaperone-subunit complex (Slonim *et al.* 1992).

Chaperones can be classed into two groups based on the length of the loop of the elongated donor strand: Those with a shorter strand belong to the FGS type, including PapD and FimC and are generally associated with rod-like pilus fibres. The other group with the longer loop on the strand belong to the FGL type, including Caf1M and Afa/Dr family and are associated with non-pilus fibres (Hung *et al.* 1996).

In FGS chaperones, the elongated donor strand is similar in length (10-20 residues) to the subunit with 4 conserved hydrophobic residues P1 to P4 and runs in parallel to the subunit (Hung *et al.* 1996). When the chaperone-subunit complex is formed the centreline of the subunit groove is curved away, hence holding it in a relatively open conformation that prevents compact packing of the subunit (Fig. 1.4). However, once the subunit-subunit complex is formed, strands in the groove move closer to each other which facilitates the formation of a tightly packed hydrophobic core. This is

achieved by the extra conserved residue from the subunit known as P5 (either a valine or leucine), which sits in the pocket to initiate the zip-in/ zip out mechanism that replaces the chaperone (Remaut *et al.* 2006; Vetsch *et al.* 2006).

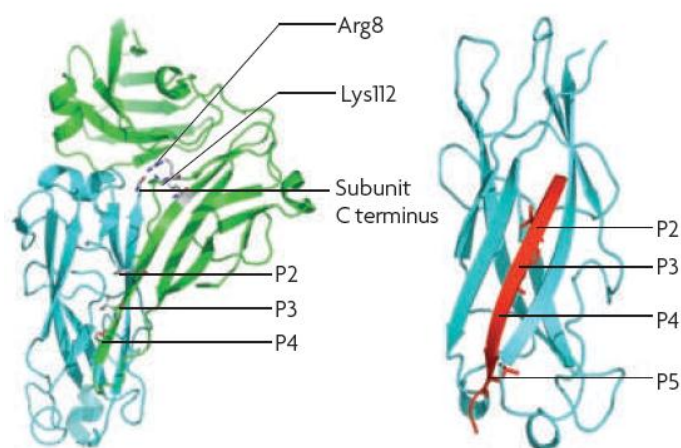


Figure 1.4: Cartoon representation of Chaperone-Subunit complex of PapD-PapK. The chaperone is coloured in green and the subunit is coloured in blue. On the left is the chaperone-subunit complex with the conserved residues mapped on the chaperone. On the right is the subunit with the donor strand from the another subunit in red with the reserved residues P2 to P5. This image is adapted from Waksman and Hultgren 2009 (Waksman and Hultgren 2009).

FGL chaperones have a relative bulky strand (21-29) which penetrates deeper into the hydrophobic core and holds the neighbouring strands in the groove distant from each other (Hung *et al.* 1996; Remaut *et al.* 2006). As a result, the subunit is in an active folding intermediate state (Sauer *et al.* 2002; Zavialov *et al.* 2003) and the stored energy is used for the fibre formation. In the subunit-subunit complex, the groove is allowed to close and promoting tight packing at the position of residue P5. The strand also releases the stored energy from the chaperone-subunit complex and transforms to a compact Ig fold domain (Sauer *et al.* 2002; Zavialov *et al.* 2003). The subunit-subunit interactions in the fibre are more stable than the chaperone-subunit interactions, as it resists dissociation in 2% SDS at room temperature (Striker *et al.* 1994; Soto *et al.* 1998).

1.4.3 Donor Strands

In the chaperone-subunit complex, the 10 to 20 residue long N-terminal extension peptide on the subunits is disordered (Sauer *et al.* 2002; Zavialov *et al.* 2003; Remaut *et al.* 2006). Despite the diversity of these subunits, the sequences of these peptides are relatively conserved (Fig. 1.5). They compete with the chaperone to promote subunit polymerisation. Conserved residues in the chaperone-subunit interface are located in the pockets P1 to P4 but in the subunit polymer the four conserved hydrophobic residues occupy pockets P2 to P5 instead (Waksman and Hultgren 2009). In particular, the P4 residue is strictly conserved because the P4 pocket is usually formed by an aromatic residue and glycine is the only residue which could accommodate it with no clashes (Waksman and Hultgren 2009).

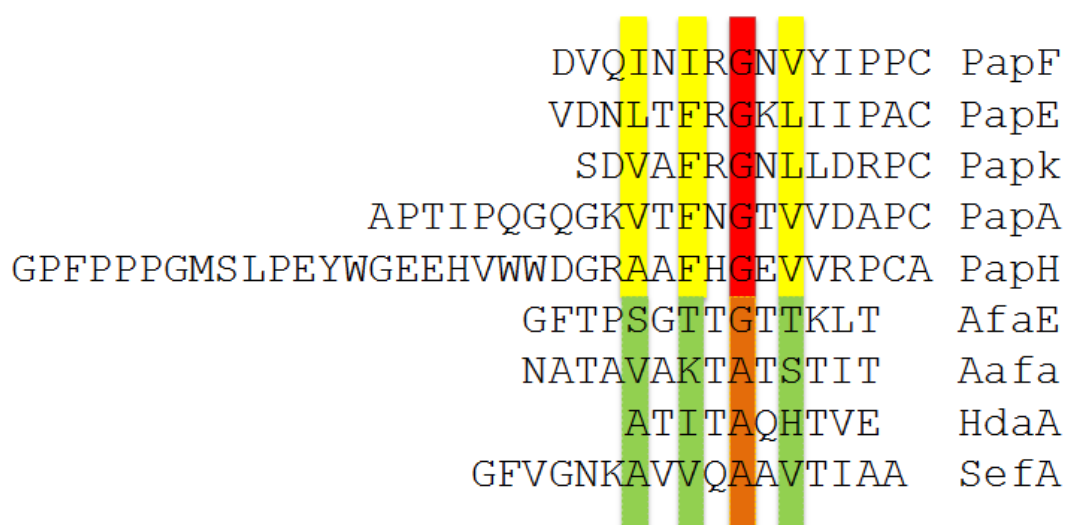


Figure 1.5: Sequence Alignment of the donor strand from Pap subunits. The coloured boxed areas are the conserved hydrophobic residues (yellow) and glycine (red box) which resident in the P4 pocket. This image is adapted from Waksman and Hultgren 2009 (Waksman and Hultgren 2009). These positions are occupied mostly by hydrophobic residues in the Afa/Dr subunits (coloured in green and orange boxes).

The subunits have an incomplete immunoglobulin-like (Ig-like) fold, which lacks a C-terminal β -strand as the result of a groove in the hydrophobic core. In the subunit-subunit complex the G1 strand from the chaperone is replaced by the N-terminal

β -strand from the adjacent subunit (zip-in / out mechanism) to complete the Ig fold, the process is known as the donor strand exchange (DSE) (Fig. 1.6) (Remaut *et al.* 2006; Waksman and Hultgren 2009).

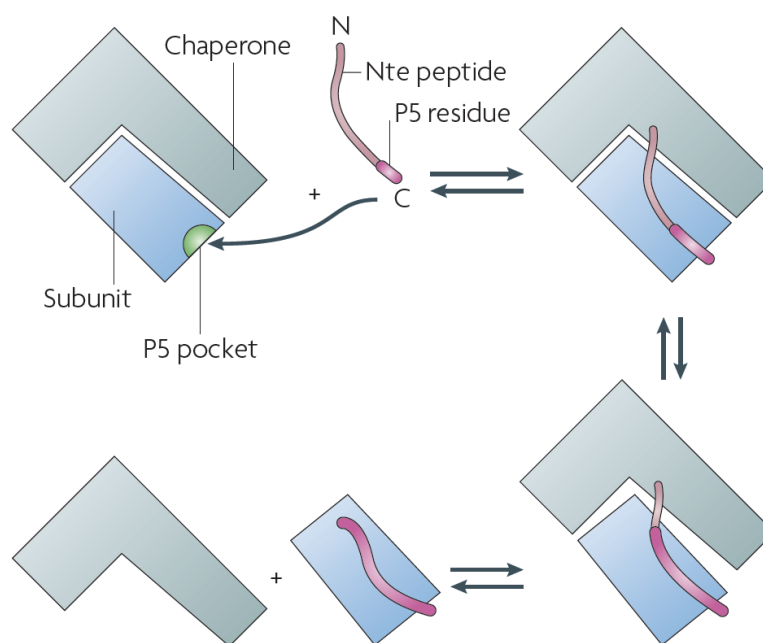


Figure 1.6: A schematic diagram for ‘zip-in, zip out’ mechanism for donor strand exchange (DSE). The residues in the subunit pockets are replaced sequentially by the insertion of the N-terminal peptide begins at position P5 (zip-in). As the residues are replaced, the chaperone releases the subunit and allows polymerisation of the subunits (zip-out). This image is adapted from Waksman and Hultgren 2009 (Waksman and Hultgren 2009).

1.5 Afa/Dr family

Afa/Dr family related adhesins can be found in various strains of pathogenic *E. coli* (i.e. DAEC and EAEC) related to diarrhoea and urinary tract infections (Le Bouguenec and Servin 2006) (Table 1.2). The *afa-I* and *dra* operons were the first to be identified (Labigne-Roussel *et al.* 1984; Nowicki *et al.* 1987). The adhesins encoded by the Afa/Dr family are transported via the chaperone-ushe system to the cell surface where they display different morphologies: afimbrial or/and fimbrial. The operons included in the Afa/Dr family share a high degree of similarity but the sequences encoding the major subunit are highly heterogeneous within the family (Le

Bouguenec *et al.* 1993; Garcia *et al.* 1994). Most adhesins of the Afa/Dr family interact with the Cromer blood group antigen on the human DAF (Nowicki *et al.* 1990; Pham *et al.* 1997). Other Afa/Dr strains (i.e. *afa-VII* and *afa-VIII*) infect animals and are unable to recognize the human DAF (Lalioui *et al.* 1999; Lalioui and Le Bouguenec 2001).

Pathotype	Afa/Dr Family	Host	Morphology
DAEC	Afa-I	Human	Afimbrial
DAEC	Afa-II	Human	Afimbrial
DAEC	Afa-III	Human	Afimbrial
DAEC	Afa-V	Human	Afimbrial
Extra-intestinal	Afa-VII	Bovine	Afimbrial
Extra-intestinal	Afa-VIII	Animal and Human	Afimbrial
DAEC	Dr (dra)	Human	Fimbrial
EAEC	Aaf-I	Human	Fimbrial
EAEC	Aaf-II	Human	Fimbrial
EAEC	Aaf-III	Human	Fimbrial

Table 1.2: Characteristics of Afa/Dr family. Afa/Dr family can infect both human and animal and shows different morphology of adhesins on the surface.

1.6 Afimbrial adhesin (Afa)

By similarity with the Pap and Fim operons, it has been established that Afa are assembled via the chaperone-usher pathway (Sauer *et al.* 1999; Soto and Hultgren 1999; Sauer *et al.* 2000; Sauer *et al.* 2000). Afa forms either an extended hair-like fibre known as a pilus (or fimbriae) or afimbrial sheaths, on the amorphous outer

membrane; these structures recognize specific receptors on host cells (Soto and Hultgren 1999).

1.6.1 Genetic Organisation

All operons encoding Afa/Dr adhesins have a similar genetic organization. For example, the *afa3* operon is 7.6 Kb and expresses six polypeptides: AfaF, AfaA, AfaB, AfaC, AfaD, and AfaE (Fig. 1.7). AfaE-III has been identified as the structural adhesin (Garcia *et al.* 1994) and is located at the same position as the adhesin on the *afa-I* operon. AfaE-III shares 99.4% identity the DraA fimbrial adhesin, and is responsible for mediating initial adhesion events directly linked to bacterial virulence (Anderson *et al.* 2004). AfaB-III and AfaC-III have been identified as a periplasmic chaperone and an outer membrane anchor protein, respectively (Garcia *et al.* 1994). AfaA-III and AfaF-III are homologous to the PapI-PapB transcriptional regulatory proteins (Garcia *et al.* 1994). The AfaD-III maps to the same locus as the invasins on the *afa-I* operon and is required for efficient internalization of the host cells.

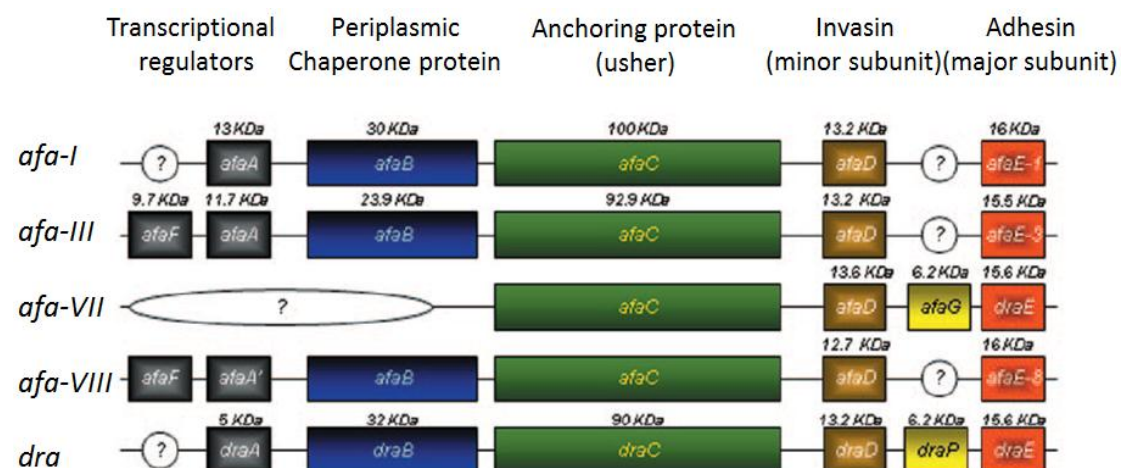


Figure 1.7: Genetic organisation of Afa/Dr operons. The same organisation is shared within the family and consists of transcription regulator, chaperone, anchoring protein (usher), invasin and adhesin (Le Bouguenec *et al.* 1993; Garcia *et al.* 1994). This image is adapted from Servin, 2005 (Servin 2005).

1.6.2 Afa structures

The structure of AfaE-III was solved by NMR whereby an additional C-terminal strand was added to mimic the subunit-chaperone complex (DSC) (Anderson *et al.* 2004). This supported the possible mechanism for the assembly of an AfaE-III polymer and producing stable monomer (Anderson *et al.* 2004) (Fig. 1.8).

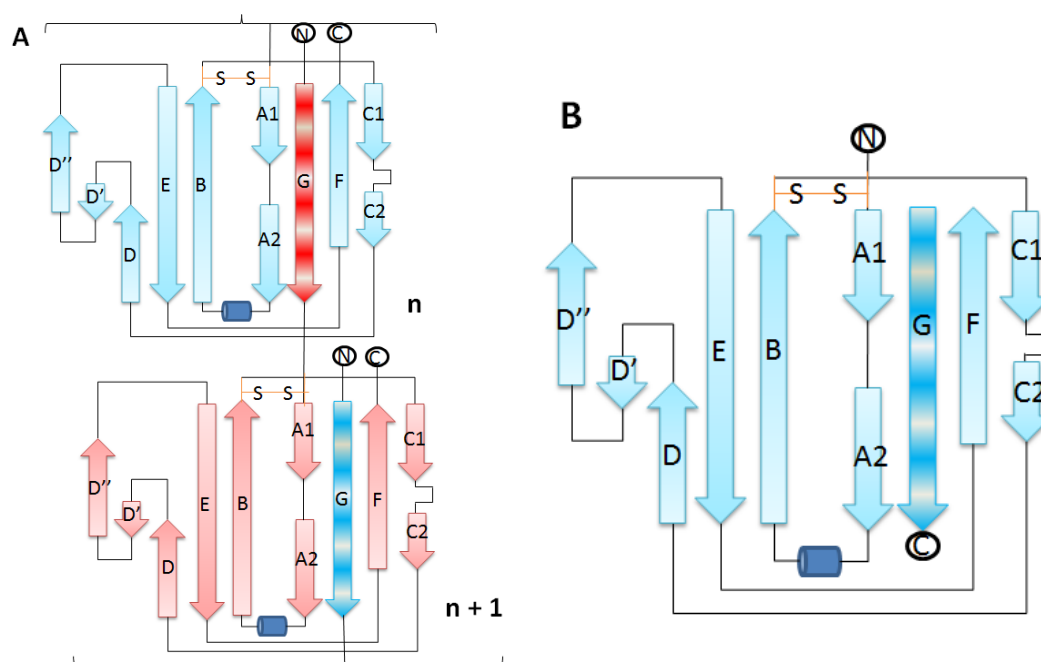


Figure 1.8: Topology diagram of atomic structure of AfaE-III. (A) Topology diagram of the DSC-induced AfaE-III polymer. (B) AfaE-Dsc which is self-complemented. This image is adapted from Anderson, *et. al.*, 2004 (Anderson *et al.* 2004).

The AfaE-III adhesins assemble into a flexible fibre that links both the bacterial membrane usher and the invasin on the tip (Anderson *et al.* 2004). An electron microscopy (EM) image of the AfaE-III polymer is shown in Fig. 1.9 (Anderson *et al.* 2004). It suggests that AfaE-III polymers form a thin (2 nm in diameter) fibre shape. However, some EM studies have observed a different polymer morphology including both thin (~2 nm in diameter) and thick fibres (Bilge *et al.* 1989; Swanson *et al.* 1991; Keller *et al.* 2002).



Figure 1.9 Images of AfaE-III filament. On the left is the EM image of Afa-III fimbrial and on the right in the schematic representation of the filament. This image is adapted from Anderson *et. al*, 2004 (Anderson *et al.* 2004).

1.6.3 Receptors

DAEC is one of the best characterised strains of pathogenic *E. coli* and many host receptors have been identified. Afa adhesins and invasins have a range of different receptors including DAF (CD55), CEACAM1 (CD66a), CEA (CD66e), CEACAM5 (CD66c) and integrin $\alpha_5\beta_1$ (Fig. 1.10).

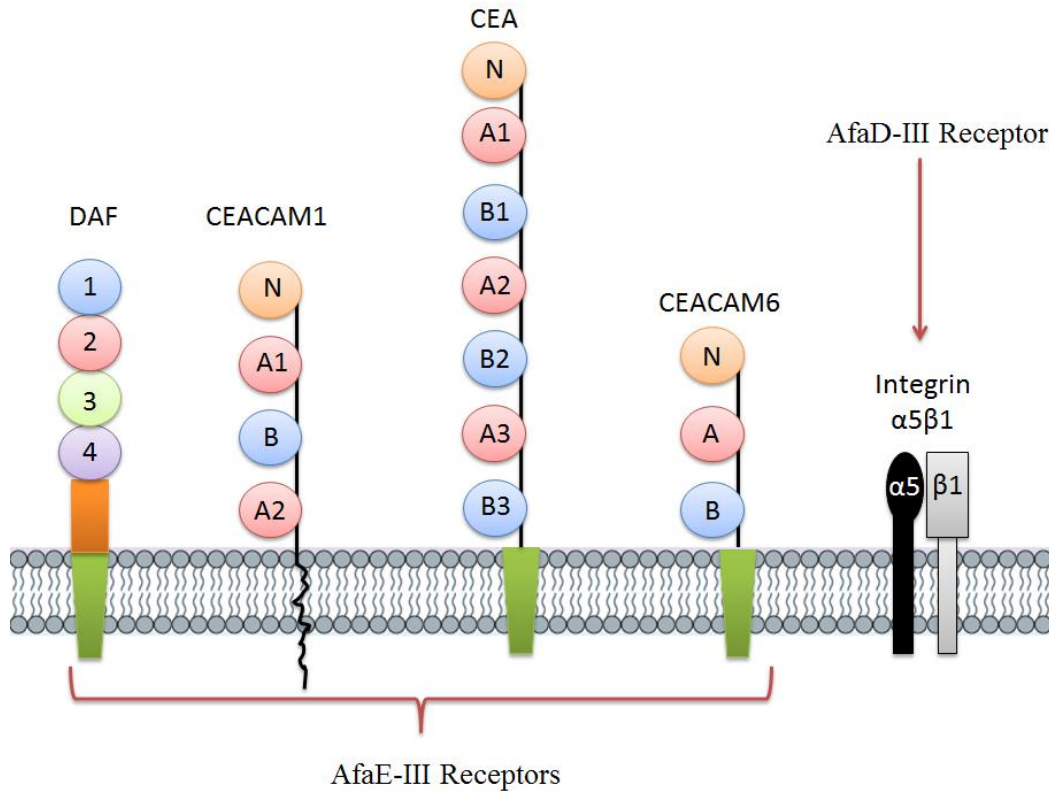


Figure 1.10: Schematic representation of different receptors of Afa/Dr adhesin and invasin. This image was adapted from Servin, 2005 (Servin 2005).

1.6.3.1 Type IV collagen and integrin $\alpha_5\beta_1$

The collagen-binding phenotype is unique to the Dr adhesin (Nowicki *et al.* 1988). The ability to bind type IV collagen is important for Afa/Dr DAEC to cause urinary tract infection and leads to the formation of persistent mesangial deposits (Miettinen *et al.* 1993). However, type IV collagen and integrin $\alpha_5\beta_1$ are components of the basement membrane involved in complex interactions at the epithelial-mesenchymal interface. Type IV collagen interacts with integrin $\alpha_5\beta_1$ expressed at the basal domain of the polarized cells (Beaulieu 1999) to form a linkage between the basement membrane and epithelial cells (Louvard *et al.* 1992). However, deregulated expression of membrane-bound molecules that are normally segregated in the basolateral domain of polarized intestinal cells occurs during inflammation and leads to a cascade of bacterial invasion.

1.6.3.2 Decay Accelerating Factor (DAF/CD55)

DAF is a complement-regulating protein (CRP) and plays a central role in preventing the escalation of the complement cascade on host cell surfaces (Servin 2005). DAF has 4 subunits, CCP1 to 4 (Fig. 1.10) and bacterial and viral pathogens recognise a variety of different subunits of this protein (Williams *et al.* 2003; Williams *et al.* 2004). All uropathogenic and diarrhoea-associated *E. coli* strains including AfaE-III adhesins recognise DAF as a receptor (Nowicki *et al.* 1988; Nowicki *et al.* 1990; Nowicki *et al.* 2001). Afa/Dr adhesins recognise CCP-3 (Nowicki *et al.* 1993) and contributes most of the free energy of binding in the interaction (Anderson *et al.* 2004).

1.6.3.3 CEACAMs

The carcinoembryonic antigen (CEACAM) family belongs to the immunoglobulin superfamily (IgSF) of adhesion molecules with a single conserved N-terminal immunoglobulin (Ig) variable (Ig_v)-like domain followed by up to 6 Ig constant (Ig_c)-like domains.

Many studies have shown that CEACAMs are one of the receptors for pathogens such as *Neisseria gonorrhoeae* and *Neisseria meningitides* (Chen and Gotschlich 1996; Virji *et al.* 1996; Gray-Owen *et al.* 1997). Once the pathogenic bacteria bind to CEACAMs, a signalling pathway is triggered which allows these pathogens to evade the host cells (Guignot *et al.* 2001). As shown in a rat model, the long cytoplasmic tails of CEACAM1 and CAECAM3 (which consist of an immunoreceptor tyrosine-based inhibitory motif and an immunoreceptor tyrosine-based activating motif, respectively), become phosphorylated and bind to protein tyrosine kinases and protein

tyrosine phosphatases leading to stimulation and termination of signalling (Guignot *et al.* 2001).

1.6.4 Mechanism of infection

The infection mechanism of Afa has been studied intensively and consists of a series of events (Fig 1.11) (Servin 2005). Initially, AfaE-III from DAEC interacts with the membrane bound receptor (DAF and CEA) on the surface of microvilli which leads to loss of microvilli as a result of signalling pathways involving protein kinases and an increase of intracellular calcium ions. This eventually depolarises the epithelial cell as the cytoplasmic and membrane-bound proteins become evenly distributed. This allows integrins, normally located on the basolateral domain, to migrate to the apical region and interact with AfaD-III leading to internalization.

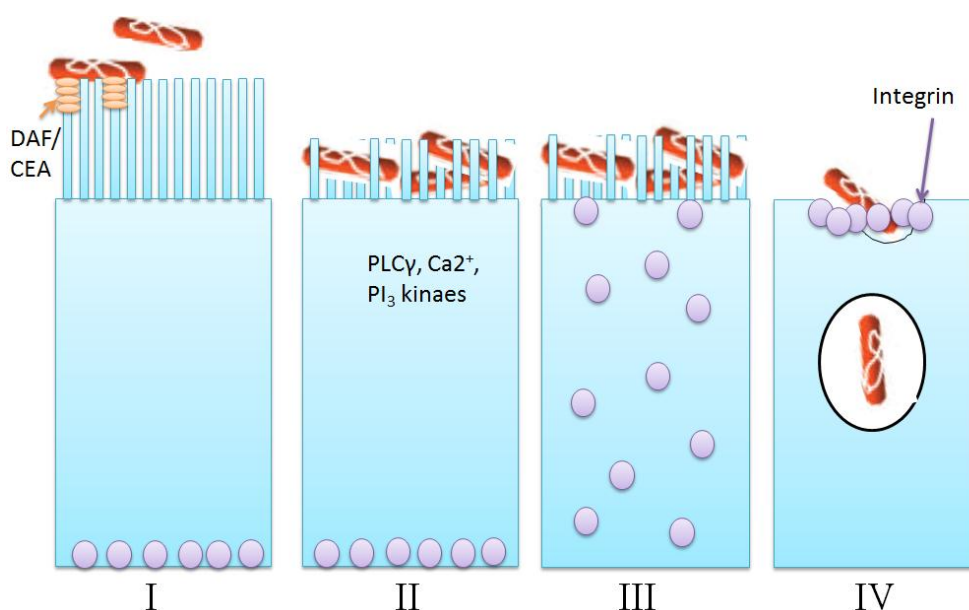


Figure 1.11: The infection mechanism of DAEC. I. Afa adhesins such as AfaE-III adhere to the receptor (i.e. DAF and CEA in orange). **II.** This leads to the loss of microvilli and triggers a cascade of signalling and an increase in intracellular calcium ions. **III.** These eventually destroy the polarization of the cell and allow proteins such as integrin (in purple) to be distributed evenly in the cell surface. **IV.** Another Afa/Dr protein such as AfaD-III can interact with integrin and induce internalisation. This image was taken from Servin, 2005 (Servin 2005).

1.6.5 Roles of Afa adhesins in infection

The mechanism of cell internalization by Afa strains has also been described. Garcia *et al.*, (2000) have shown that strains bearing the *afa-III* operon can trigger cell invasion at reduced levels in the absence of the AfaD invasin. Similarly, it has been shown that binding of the Dr adhesin is able to promote cell invasion via the short consensus repeat 3 domain of the GPI anchored decay accelerating factor, DAF (Selvarangan *et al.* 2000). The dissociation constant of this interaction has been characterized and falls within the micromolar range (16 μ M) (Anderson *et al.* 2004).

Several studies have suggested AfaE is the main and possibly only essential protein for invasion. In 1997 *afaD*⁻ and *afaE*⁻ strains were used to infect HeLa cells (Jouve *et al.* 1997). The results show that the *afaD*⁻ strain exhibits a diffuse adherence pattern similar to the strain with a wild type *afa* operon but lower levels of invasion, whereas the *afaE*⁻ strain fails to either associate with or invade the epithelial cells. Furthermore, it has been shown that infection of HeLa cells by *E. coli* (DH5 α) expressing Dr, Dr-II, F1845, Afa-I and Afa-III adhesins, ultimately leads to internalization (Goluszko *et al.* 1999). In addition, a Dr⁺ *E. coli* (BN406) strain is internalized by CHO DAF⁺ cells, presumably via interaction with human DAF as an AfaE receptor (Selvarangan *et al.* 2000). Together these results suggest that AfaD is non-functional for either adhesion or invasion especially without the presence of AfaE. However, as the mechanism of pathogenicity and the known receptors suggest, AfaE interacts with the surface receptors (i.e. DAF and CEA) whereas AfaD can be functional as an invasin via recognition of integrin (basolateral protein) in depolarised cells.

In addition, an alternative mechanism for promoting entry of Afa/Dr DAEC into epithelial cells has been demonstrated (Guignot *et al.* 2001). *E. coli* 1H11128 (clinical

isolate) can efficiently invade HeLa cells in the absence of DAF (CD55) and CEA (CD66e), both receptors of AfaE. This indicates that another basolateral domain-associated molecule can function as a receptor for IH11128 cell entry, which is known to be triggered by the interaction between AfaD and the $\alpha 5 \beta 1$ integrin. Disruption of the cell to cell junctions is carried out by the pathway involving caveolae and dynamic microtubules. A cascade of events begins with the disruption of cell to cell junctions which causes the epithelial cells to lose their cell polarization. This leads to the diffusion of integrins to the apical domain where they can interact with AfaD. Crucially, Guignot *et al* have confirmed that the interaction of integrin with AfaD-III can lead to a cascade of internalization events and have also discovered the secreted autotransporter toxin, Sat (within the subfamily of serine protease autotransporters of Enterobacteriaceae, SPATEs), functions as a virulence factor in Afa/Dr DAEC by promoting lesions in tight junctions of polarized epithelial cells (Guignot *et al.* 2007). Sat is responsible for the Afa/Dr DAEC-induced lesions in the tight junctions of polarized intestinal cells. In monolayer-forming cells, Sat also induces and increases paracellular permeability and the rearrangement of tight junction-associated proteins without transepithelial resistance modification (Guignot *et al.* 2007). This establishes an AfaE-independent mechanism of invasion.

In studies that favour an AfaE dependent mechanism of invasion, they have all over-expressed either DAF or CEA, which could overstate the contribution of AfaE in the adhesion/invasion mechanism. On the other hand, the studies which favour a functional role for AfaD as an invasion provide a new perspective on the possibility that AfaE and AfaD act synergistically in the adhesion/invasion mechanism. Therefore, different homologues from the Afa/Dr family need to be investigated to reveal the true importance and function of AfaD.

1.7 Aggregative Adherence Fimbriae (Aaf)

A flexible, bundle-forming fimbrial structure with sizes varying from 2-3 nm to 6-7 nm in diameter, was found in clinically isolated EAEC strains (Knutton *et al.* 1992) and is encoded by the *agg* genes. These fimbriae can be distinguished into three classes, Aaf-I, Aaf-II and Aaf-III. Aaf-I are organized as two different gene clusters separated by a 9 Kb DNA (Nataro *et al.* 1993). Region 1 contains a cluster of genes required for fimbrial synthesis and assembly, comprising the structural subunit of the fimbriae, *aggA* (Savarino *et al.* 1994), a member of the Afa/Dr Family of adhesins. Region 2 is comprised of *AggR*, a transcriptional activator of aggregative adherence fimbriae I (Aaf-I) expression (Savarino *et al.* 1994). However, only a minority of EAEC strains express Aaf-I (Nataro *et al.* 1993; Nataro *et al.* 1994), Czeczulin *et al.* have found on a second plasmid coded for a 5 nm diameter fimbria (Aaf-II) (Czeczulin *et al.* 1997) that is morphologically, antigenically and genetically distinct from Aaf-I. Aaf-II is organised as two unlinked regions, region 1 includes *aafA*, the structural fimbrial subunit, *aafR* and the chaperone *aafD*, whereas the usher protein *aafC* is located within region 2. Aaf-II required expression of the transcriptional activator *aggR* and *fis* gene is required for the expression of *aggR*, which in turn encodes an activator for *aafD* (Elias *et al.* 1999; Sheikh *et al.* 2001).

Furthermore, a different aggregative adhesive fimbria (Aaf-III) has been found in a typical EAEC strain (Bernier *et al.* 2002) which is closely related to *agg* and *aaf* and also required the activation of the transcriptional activator *aggR*.

1.7.1 Gene Cluster of Aaf-II

Aaf-II cluster is separated into two regions by enterotoxins Pet, EAST1 (*astA*) (Fig. 1.12). Region 1 consists of *aafD*, *aafA*, insertion-like 1 (*IS1*) and *aggR* genes. *aafD* is

747 bp in length and encodes a typical Sec-dependent signal sequence and homologous to the chaperone gene *aggD* of the Aaf-I cluster, with 62.4% sequence identity and 88.7% similarity. The *aafA* gene product which encodes the fimbrial subunit is homologous to the *aggA* gene product of the Aaf-I at 25% identity (Czeczulin *et al.* 1997). Downstream of *aafA* is the IS1 element which has 98% identity to the IS1 from *afa-3* cluster. The *aggR* gene is 795 bp in length with 99% identical sequence to the *aggR* from the Aaf-I and belongs to the AraC family of transcriptional activators family (Gallegos *et al.* 1997). Studies have shown that clones expressing only the *aafD-aafA-aggR* cluster do not give fimbrial express fimbria, which suggest the need of cluster 2.

Cluster 2 consists of *aafB*, *aafC*, *aafD'* and IS1. The *aafB* is 438 bp in length which includes a cleavable signal sequence and the gene product has 59% and 56% identity to the putative invasion protein AfaD (Afa/Dr family of adhesins) and AafB of the Aaf-I cluster, respectively (Jouve *et al.* 1997). The predicted *aafC* gene is 2,583 bp in length and has 67% identity to *aggC* (the usher protein of the Aaf-I cluster) and has 80% identity at the amino acid level to the usher protein of Afa-III (Elias *et al.* 1999). Upstream of *aafC* is the *aafD'* gene which features several frame shifts and should not generate a functional chaperone protein as the predicted mRNA suggested, but still has 57% identity to the *aafD* of *aaf-I* and 61.2% identity to *afaB*, which is required for expression of Afa-III afimbrial adhesins (Elias *et al.* 1999).

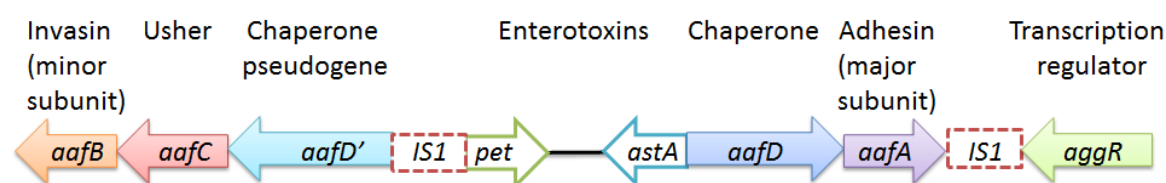


Figure 1.12: Gene cluster of Aaf-II. The Aaf-II gene cluster is separated by two enterotoxin genes, *pet* and *astA*.

1.7.2 Receptors

AafA and AafB are members of the Afa/Dr family and display high sequence similarity to AfaE and AfaD respectively but no studies have reported binding to known receptors of Afa/Dr adhesins. Recently, studies have demonstrated an interaction between extracellular matrix proteins and the major subunit of Aaf-II, AafA (Farfan *et al.* 2008).

1.7.3 Mechanism of infection

The mechanism of infection of EAEC has not yet been established and with an increasing recognition that it is usually the causative agent in persistent diarrhoea in children and adults, there is much research being undertaken on EAEC. It is believed that EAEC generates a biofilm to enhance colonization and allows it to release toxins which facilitate in the penetration into host cells (Fig 1.1) (Kaper *et al.* 2004). A model has been proposed which involved three stages (Nataro and Kaper 1998). Stage 1 involves the initial contact to the intestinal mucosa and or the mucus layer by Aaf-I and Aaf-II, adhesins proposed to be responsible for the adherence of the *E. coli* strain. During Stage 2 a biofilm is formed to facilitate persistent colonization, increase immune avoidance and perhaps increase nutrient malabsorption. In stage 3, as a result of irreversible damage to the host cell due to release of cytotoxins and enterotoxins, the *E. coli* strain is allowed to penetrate its host.

1.8 HUS-Associated Diffuse Adherent (Hda)

As previously described, Aaf family has three members but recently an uncharacterised variant named Hda which shares the same regulator as in the Aaf family, AggR, and the same aggregative adherent characteristic (Boisen *et al.* 2008).

Hda exhibits higher similarity to Afa-VIII than the rest of the Aaf family. Hda and Afa-VIII both lack of N-terminal disulphide bond presence in other members of the Aaf family which might be vital for structure or binding to the receptor (Boisen *et al.* 2008).

The Hda operon consists of 4 genes *hdaD*, *hdaC*, *hdaB* and *hdaA* (Fig.1.13). *hdaD* and *hdaC* encode for the chaperone and usher proteins respectively. The Hda fimbriae are built by two subunits *hdaA* and *hdaB* and the overall architecture is predicted to be similar to the Afa/Dr/Aaf adhesins.

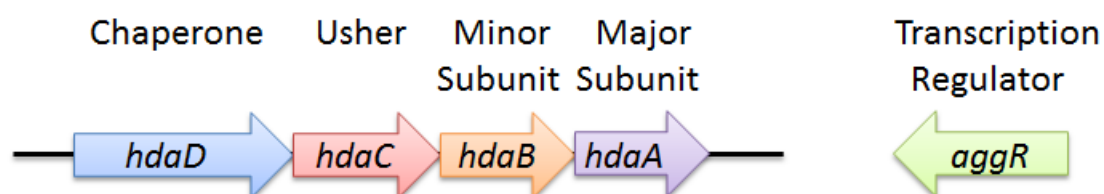


Figure 1.13: Genetic organisation of the Hda operon.

1.9 *Salmonella*

Salmonella is closely related to *E. coli*, it is a rod-shaped gram-negative enterobacterium with the diameter around 0.7 to 1.5 μm and lengths from 2 to 5 μm . It can be found in both cold and warm-blooded animals including humans and cause typhoid fever, paratyphoid fever and foodborne illness. It can survive as a free living organism for weeks and cannot be destroyed by freezing. A common method to kill *Salmonella* is by ultraviolet radiation and heat (55 $^{\circ}\text{C}$) for one hour.

Initially *Salmonella* species were divided and named according to clinical symptoms but later it was realised host specificity does not exist so it is divided into two species, *S. enterica* and *S. bongori* and 6 subspecies of *S. enterica*: *S. enterica enterica*, *S.*

enterica salamae, *S. enterica arizonae*, *S. enterica diarizonae*, *S. enterica houtenae* and *S. enterica indica*.

1.9.1 Food Poisoning from *Salmonella*

Salmonella causes food poisoning and *S. enteritidis* (a serovar of *S. enterica enterica*) is one of the leading causes of salmonellosis in humans in both North America and Europe (Rampling *et al.* 1989; Todd 1990; Goodnough and Johnson 1991; Pohl *et al.* 1991; Rivera *et al.* 1991). In healthy adults, infection requires a large amount of bacteria to overcome the acidity in stomach. However, a much smaller number is needed to cause illness in infants and young children and bacteria can also enter through inhalation of bacteria-laden dust. The incubation time is just a few hours to one day, leading to intestinal inflammation with diarrhoea and dehydration in infants. The condition can be particularly serious for weakened elderly patients.

1.10 *Salmonella enteritidis* Fimbriae (Sef)

Three distinct fimbrial proteins have been identified, Sef14, Sef17 and Sef21 for *Salmonella*, a closely related gram-negative bacterium to *E. coli* (Feutrier *et al.* 1986; Feutrier *et al.* 1988; Collinson *et al.* 1991; Muller *et al.* 1991). Among the many strains of *Salmonella*, Sef14 is restricted to *S. enteritidis* and other closely related group D *Salmonella* (Clouthier *et al.* 1993). A *sef* operon is located on a small pathogenicity island which is required for the translation and biogenesis of Sef14 fimbriae. This operon exhibits a similar chaperone-usheer system to Afa/Dr/Aaf family.

1.10.1 Biogenesis of *sef* genes

The Sef operon consists of 4 genes, *sefA*, *sefB*, *sefC* and *sefD* and is regulated by a downstream AraC-like regulator *sefR* (Lindler and Tall 1993). The chaperone protein

is encoded by *sefB* and *sefC* is the usher protein. Fimbriae are built by two subunits SefA and SefD, the major subunit and minor subunit respectively. SefD is essential for the export of SefA but SefD can be observed on the surface without the presence of SefA (Edwards *et al.* 2000) (Fig. 1.14). In addition, without SefA, *S. enteritidis* strain remains virulent and so it is likely to position at the tip of the fimbriae and could be an adhesin for host recognition. (Edwards *et al.* 2000).

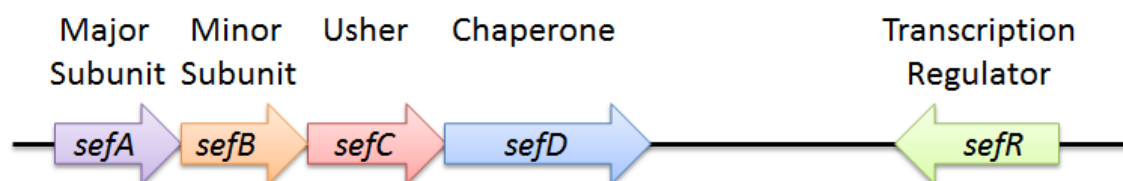


Figure 1.14: Operon of Sef and proposed mechanism of fimbria formation. The Sef operon consists 4 genes and is regulated by downstream *SefR*.

1.10.2 Role of Infection

It has been shown that Sef may not be involved in the initial interaction with the epithelia cell (Dibb-Fuller *et al.* 1999) despite the fimbriae formation (Ogunniyi *et al.* 1994; Ogunniyi *et al.* 1997). Instead, it has been suggested that Sef is essential for efficient interaction with macrophages, which lead to colonization and penetration through the intestinal barrier (Edwards *et al.* 2000). Sef may be required for the initial contact with the macrophage where it can interact with a receptor on the surface to alter the uptake into the phagocyte and survive the intracellular stage (Edwards *et al.* 2000). In addition, SefD is thought to be essential for *S. enteritidis* to survive within active macrophages but no clear mechanism has been found.

1.11 Project Aims

The true function of AfaD remained controversial and in this project, different proteins from Afa/Dr family would be investigated and compared structurally to

identify potential receptors. As the major subunit, AfaE-like proteins are more diverse therefore AfaD-III was used to identify homologous proteins and sequence alignment (Appendix 1). The phylogram was produced based on this alignment (Fig 1.15).

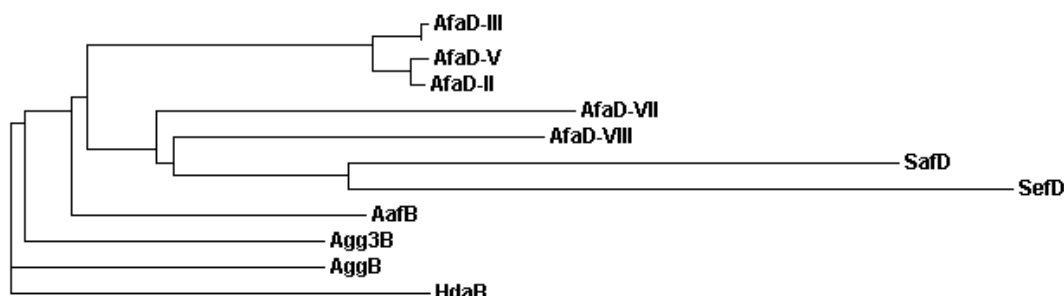


Figure 1.15: Phylogram based on the alignment of AfaD-III homologues. The phylogram was produced using ClustalW and phylogeny from EBI.

Proteins from different branches such as AafB, HdaB, SefD, Agg3B and AfaD-VII were used to investigate their structural and functional characteristics, as they can provide more information on how the diversity of the Afa/Dr family and other strains of *E. coli* and *Salmonella* related to variation infection mechanisms.

This project aims are as follows:

- Genetically engineer donor strand complemented constructs of AfaD-VII, AfaE-VII, AfaDE-VII, AafB, AafBA, Agg3A, Agg3B, Agg3BA, HdaB, HdaA, HdaBA, SefD, SefA and SefDA.
- Test expression for all constructs.
- Scale up and optimizing expression conditions to maximise protein yield.
- Purifying proteins for structure determination.
- Elucidate structures using appropriate structural technique (X-ray crystallography or NMR).
- Carry out functional studies by investigating potential binding sites.

Chapter 2: Theories of Structural Techniques

2.1 Introduction

Over the last century, the advancement of biophysical techniques to elucidate the structures of biological molecules has pushed forward the scientific understanding of how molecules recognise each other and function at the atomic level. For example, structural information has provided great insight into the donor strand exchange mechanism in the chaperone-usheer system and has facilitates the design and production of stable fimbrial proteins (Striker *et al.* 1994; Choudhury *et al.* 1999; Sauer *et al.* 1999).

A number of different complementary techniques are available to determine the structure of a protein. The quaternary structure of large protein oligomers and complexes can be easily determined by electron microscopy, although the resolution is generally rather low. In order to obtain the high resolution information needed to resolve the atomic coordinates of a macromolecule, nuclear magnetic resonance (NMR) and X-ray crystallography are generally used. Both X-ray crystallography and NMR have been used in this project for structure determination and interaction studies, respectively, therefore some basic theory is described in this chapter (Blundell and Johnson 1976; Blow 2002; Keeler 2005; Tilley 2006; Cavanagh 2007; Claridge 2009).

2.2 X-ray Crystallography

X-ray crystallography is a widely used method to determine the arrangement of atoms within a crystal, where a beam of X-rays strikes a single crystal and produces a diffraction pattern. From the relative phases and intensities of these diffracted beams, a three-dimensional picture of the electron density is produced and the positions of the atoms in the crystal can be determined. However, crystal structures of proteins were not possible until 1960, when the first protein structure of myoglobin was determined

(Kendrew 1962). Since then X-ray crystallography has been developed into a powerful tool and is routinely used by structural biologists.

2.2.1 X-rays

2.2.1.1 Properties of X-rays

Electromagnetic waves can be described as disturbances of electrical and magnetic fields and have a huge spectrum with the wavelength range from several kilometres (long-wave radio waves) down to γ -rays with wavelengths in picometers (Fig. 2.1). X-rays with a wavelength of $2 \text{ \AA}$ are known as “soft” X-rays which are absorbed easily by air and water, but X-rays with shorter wavelength can be produced ($0.5 \text{ \AA}$ to $1.6 \text{ \AA}$), comparable to the distance between two atoms, to study matters at atomic level.

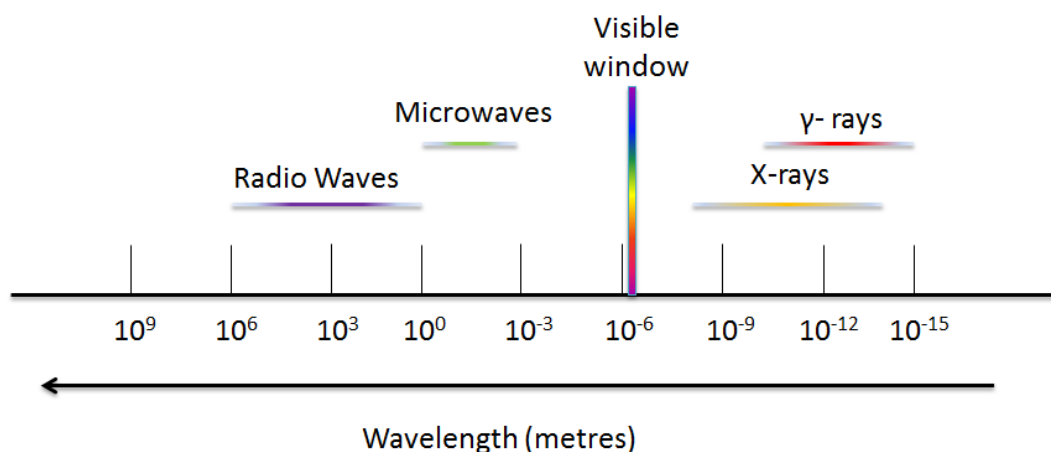


Figure 2.1: Electromagnetic Spectrum from radio waves to γ -rays.

X-rays exist as photons which exhibit both particle-like and wave-like properties. The energy carried in one photon is related to its wavelength by Planck’s equation (Equ. 2.1). Photons in the $0.5 \text{ \AA}$ to $2.0 \text{ \AA}$ range (X-rays) have energies from 25,000 to 6,000 eV which are more than sufficient to ionize atoms to a free radical form and therefore

X-rays are classified as “ionizing radiation”. Structure determination is possible by measuring the coherent scattering which occurs when the excited electrons emit the same energy and wavelength of the original X-ray that was absorbed. However, radiation damage is inevitable as the result of an electron emitting one or more photons of lower energy and remaining in an excited state which is known as incoherent scattering.

$$E=h\nu$$

Equation 2.1: Planck’s formula. E is expressed as energy with ν as frequency and h as Planck’s constant.

The X-rays used in crystallography are not very penetrating and depending on the wavelength used, the ideal size for a protein crystal is from 200 to 1,000 μm , giving a specimen which scatters a great deal of the incident beam, but does not absorb so much as to lose the signal. In addition, higher resolution can be gained and smaller crystals used with more intense, focused X-rays such as produced in synchrotrons.

2.2.2 Crystals

Crystals present a protein’s most highly ordered form and provide the special possibility of imaging the molecules. In an ideal situation all the units that form the crystal are identical and in an identical environment.

2.2.2.1 Crystallization of Biological Molecules

Molecular structure determination by X-ray crystallography is dependent on first obtaining a diffraction pattern focusing an X-ray beam onto a regular, repeating array of identical molecules (as in a crystal). The repeating unit forming the crystal is known as the unit cell and contains one or more molecules. Unlike crystals formed

from small molecules, protein crystals are generally much more delicate and do not diffract as readily to the same order. This is due to inherent internal disorder caused by the packing of irregular surfaces and the often large solvent channels which run throughout the crystal.

One of the largest bottlenecks in X-ray crystallography is obtaining high quality protein crystals which diffract. The crystallization timescale is on the range of few hours to over a year and this process is heavily dependent on factors such as pH, temperature, protein purity, protein concentration, the nature of the solvent and precipitant and the presence of ions and ligands. As the crystallization condition is hard to predict for each protein of interest, commercial kits containing vast numbers of various crystallization conditions are available and are useful for the initial screening of new protein samples. In addition, for a successful crystallization of a protein, the purity of the protein sample is also critical; it should be ideally more than 97%.

One of the most common methods for growing protein crystals is the hanging-drop method of vapour diffusion where a drop of protein solution is mixed with an equal volume of well solution (Fig. 2.2). This is suspended, within a sealed environment, above a much larger volume of well solution and once equilibrium between the well and the droplet is reached by vapour diffusion, nucleation and crystal growth may take place.

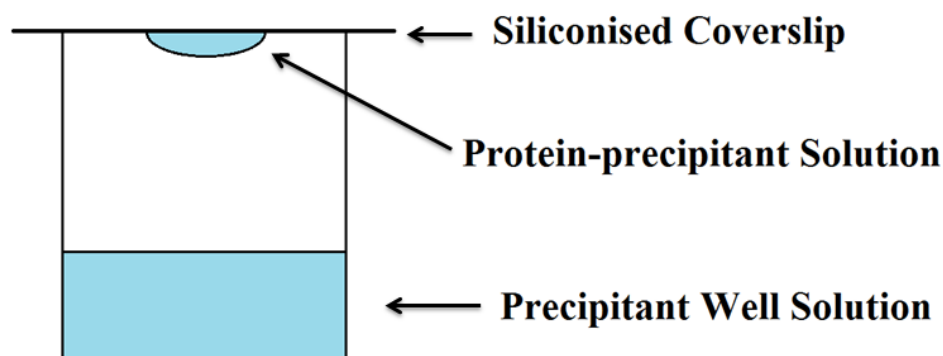


Figure 2.2: An illustration of the hanging-drop method of vapour diffusion.

This setting exhibits a diffusion gradient between the drop and the well solution as the concentration of precipitant is higher in the well solution. This diffusion promotes water molecules to leave the protein drop, which becomes supersaturated. Under the right conditions, protein molecules will aggregate and form nucleation points to allow the growth of crystals, known as the labile phase (Fig. 2.3). As the protein begins to aggregate, the concentration of the protein in the drop will decrease and enter the metastable phase where the crystals would start growing (McPherson, 1990). In order to find the appropriate conditions for crystallisation, a series of screens of different solutions with different pH, buffer and precipitants are used. Once the conditions have been discovered, these can be optimized by making small changes to the initial conditions to find the optimal condition for crystallisation and yield a high quality diffracting crystal.

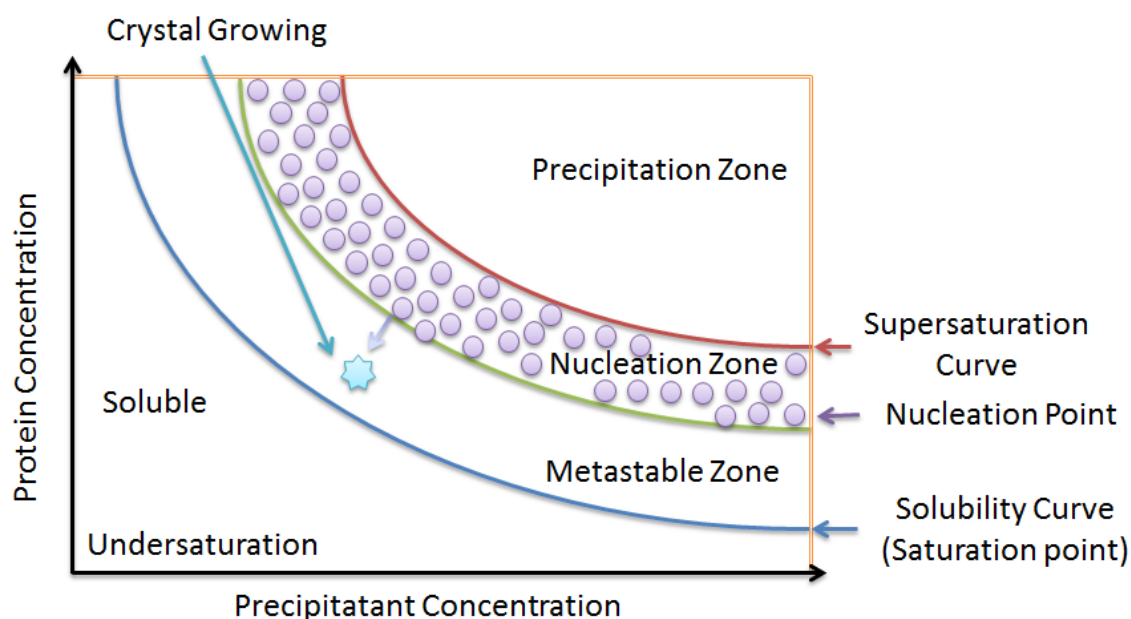


Figure 2.3: Phase diagram illustrating protein solubility as a function of precipitant concentration. This diagram represents a hypothetical protein, with the red line showing the maximum solubility of the protein. The green line represents the region between the metastable and nucleation zones. Nucleation points formed in the nucleation (labile) zone and as protein concentration reduces so the nucleation points falls into the metastable (labile) zone where crystal (blue star) is growing.

2.2.2.2 Crystal Lattice and Symmetry

Crystals represent matter in its most highly ordered form. They are built up in a regular way so that equivalent units each interact identically with their neighbours. The way the molecules stack together allows a crystal to have flat faces with sharp edges. A perfect crystal has lattice symmetry where any one of the repeating units of its structure can be placed in the exact position and remain indistinguishable. These repeating units, defined in three-dimensions are known as unit cell and are the fundamental building blocks whose edges correspond to the three lattice translations (Fig. 2.4). Positions in the unit cell are defined by distances x , y and z representing fractional displacements along these three axial directions. It is usual to define a unit cell which has the smallest possible volume. It is the smallest unit which can build up the whole crystal by lattice symmetry operations. The three lattice translations

defining a unit cell are called **a**, **b** and **c** and are commonly understood as the right-handed set.

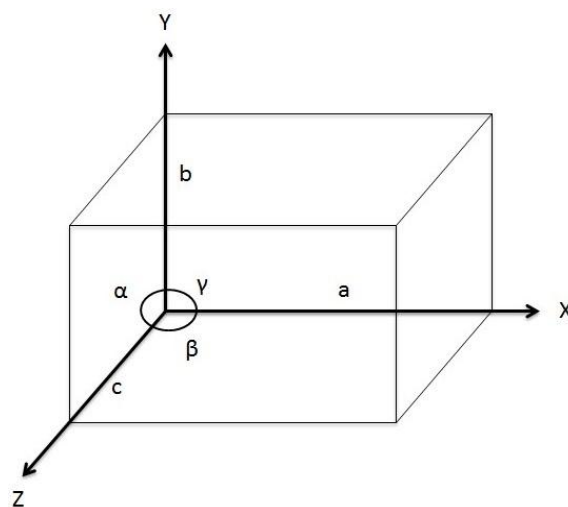


Figure 2.4: An example of a unit cell from a crystal lattice. The vectors **a**, **b** and **c** describe the length of the unit cell on the **x**, **y** and **z** axis. The angles in the unit cell are described by α , β and γ .

Lattices can have the same type of symmetry as any objects, but rotational symmetry applies restrictions on the shape of the unit cell. The stack of the unit cells must fill space and fit the rotation therefore the lattice translations have to fulfil the symmetry as well. For example, in the case of two-fold symmetry, there must be two lattice translations which are perpendicular to the symmetry axis. Three-fold, four-fold and six-fold symmetry also have similar requirements as one lattice translation must be parallel to the rotation axis, and perpendicular to two other lattice translations. It is not possible to fill a plane with regular pentagons therefore no lattice with 5-fold symmetry can exist. In some cases the perpendicular directions are chosen for the lattice translations, defining the conventional unit cells, which then contain more than one lattice point. This means the lattice may be found to be face-centred or body centred which means the lattice points are located at the corner of the cell. Furthermore, depending on the position of the symmetry lattice points either in the

centre of one face, at the centre of all the faces, or at the centre of the cell, four types of lattice centring can be obtained (Fig 2.5): primitive centring (P), body centred (I), all face centring (F) and single face centring (A, B or C). Primitive centring is the simplest arrangement with one lattice point in each of the cell corners and body centred has an additional lattice point at the centre of the cell. All-face centring has the lattice points in all corners as well as one additional lattice point at the centre of each face of the cell, whilst with single face centring the additional lattice point is just on a single face.

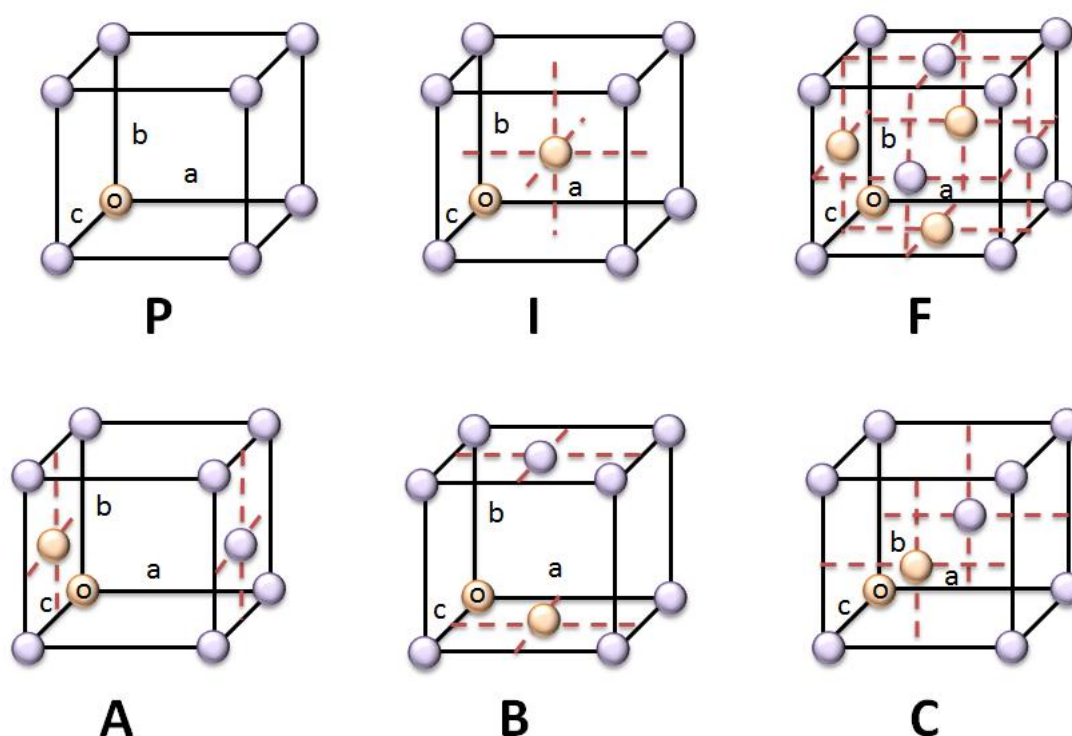


Figure 2.5: Centring arrangement within a Unit cell. Black lines represent the the unit cells and the edges are labelled, a, b and c. The orange spheres are unique lattice points, with the purple spheres correlate to the related lattice points. The red dotted lines show the centering inside the unit cell and the face centering of lattice points. P = primitive unit cell, I = body centered unit, F = all face centered unit cell, A, B and C = unit cell center on the a (101), b (010) and c (001) planes.

The combinations of these symmetry operations give seven crystal systems and are named according to the relationship between the axes and the inter-axial angles (Table 2.1). According to the symmetry of the crystal lattice, the conventional unit cell often

contains several equivalent units, some of them rotated relative to others by symmetry operations. The whole symmetry of the crystalline lattice is defined by its space group. This is the complete group of crystal symmetry operations that generate the three dimensional lattice and define its symmetry, and other operations that relate different units within the cell. The smallest unit of the crystal that can generate the complete crystal structure is known as the crystal asymmetric unit. There are 42 combinations of lattice centring but many of them are equivalent which reduces to only 14 unique three dimensional lattices known as the Bravais lattices or direct lattices.

Crystal System	Minimum Symmetry Requirement	Constraints on Interaxial angles and axial lengths
Triclinic	None	$a \neq b \neq c; \alpha \neq \beta \neq \gamma$
Monoclinic	Two-fold axis on b or c	$a \neq b \neq c; \alpha = \gamma = 90^\circ, \beta \neq 90^\circ$ (b is the unique axis) or; $\alpha = \beta = 90^\circ, \gamma \neq 90^\circ$ (c is the unique axis)
Orthorhombic	Three perpendicular two-fold axis	$a \neq b \neq c; \alpha = \beta = \gamma = 90^\circ$
Tetragonal	Four-fold axis	$a = b \neq c; \alpha = \beta = \gamma = 90^\circ$
Trigonal	Three-fold axis	$a = b \neq c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$ (hexagonal axis) or; $a = b = c; \alpha = \beta = \gamma \neq 90^\circ$ (rhombohedral axes)
Hexagonal	Six-fold axis	$a = b \neq c; \alpha = \beta = 90^\circ; \gamma = 120^\circ$
Cubic	Four three-fold axis	$a = b = c; \alpha = \beta = \gamma = 90^\circ$

Table 2.1: Definition the seven crystal systems.

The internal planes through a crystal structure are specified in terms of Miller Indices, h , k and l which represent one plane and all other identical parallel lattice planes. The

values of h , k and l are the reciprocals of the fractions of a unit cell edge, a , b and c respectively, intersected by an appropriate plane (Fig.2.6).

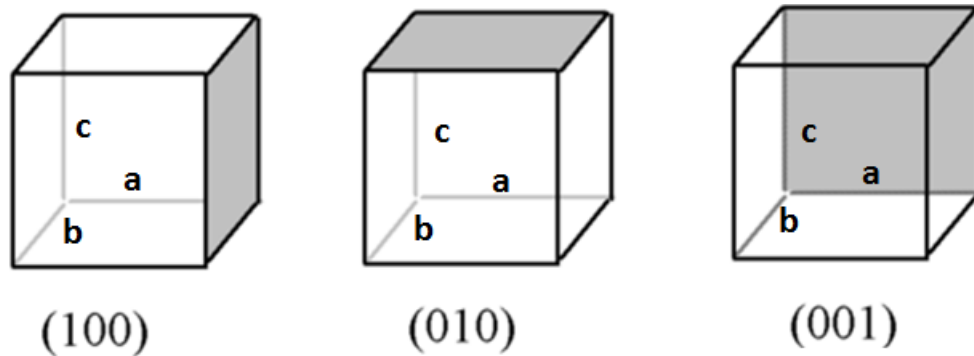


Figure 2.6: Example of miller indices.

2.2.3 X-ray Diffraction by a Crystal

X-rays interact with free electrons and alter the energy state. As these electrons return to their original energy state, X-ray photons are emitted in all directions which are equivalent to the differences in energy. In 1912, Max von Laue discovered that this coherent scattering can be detected and will only occur when the two atoms in a lattice translation are separated by an integral numbers of wavelengths. Furthermore, Lawrence Bragg in 1913 described the angles needed for coherent and incoherent scattering from a crystal lattice and showed that diffraction can be detected when the scattered waves interfere constructively and would only occur when the path length of each wave is equal to an integer multiple of the wavelength (Equ. 2.1 and Fig. 2.7) .

$$n\lambda = 2d_{hkl} \sin\theta$$

Equation 2.1: Bragg's Law. n is an integer, λ is the wavelength of the X-rays, d_{hkl} is the spacing between planes, θ is the angle of diffraction

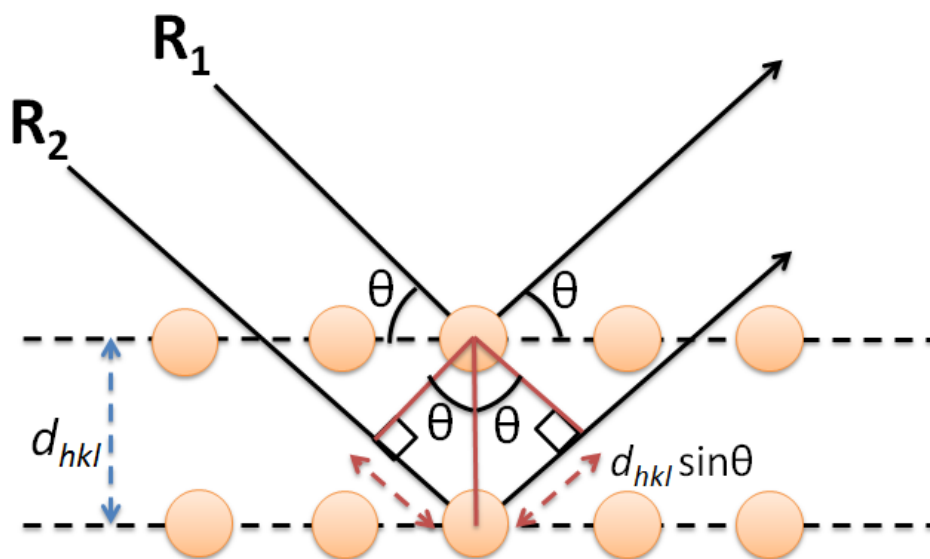


Figure 2.7: Diagram representing Bragg's law. X-ray beams R_1 and R_2 impose on atoms (orange spheres) within two parallel planes (dotted black line) of a crystal lattice with an angle θ , separated by distance d_{hkl} . The extra length travelled by X-ray beam R_2 ($2d_{hkl} \sin\theta$) can be determined by using the sine rule.

A diffraction pattern is observed and collected in X-ray crystallographic experiments and is related to the position of atoms within each set of parallel planes with coordinates of hkl . The distance between these reflections are the reciprocal lengths of the unit cell vectors a , b and c . Therefore, the dimensions (a^* , b^* , c^*) are related reciprocally to the real lattice dimensions of the crystal. A perpendicular line of length $1/d_{hkl}$ from the real space lattice can be drawn to construct the reciprocal lattice and share the same origin. Since each reciprocal lattice point is dependent on the real space plane when the crystal is rotated the reciprocal lattice will also rotate in the same fashion.

In addition to Bragg's law, the Ewald sphere also illustrates this theory in reciprocal space (Fig. 2.8). The sphere has a radius of $1/\lambda$ and C is the point where X-rays passing through the centre of the sphere exit via the origin of the real and reciprocal lattice (O). As the crystal rotates the reciprocal lattice point will come into contact with the surface of the sphere (P), Bragg's law is satisfied, and reflection is observed.

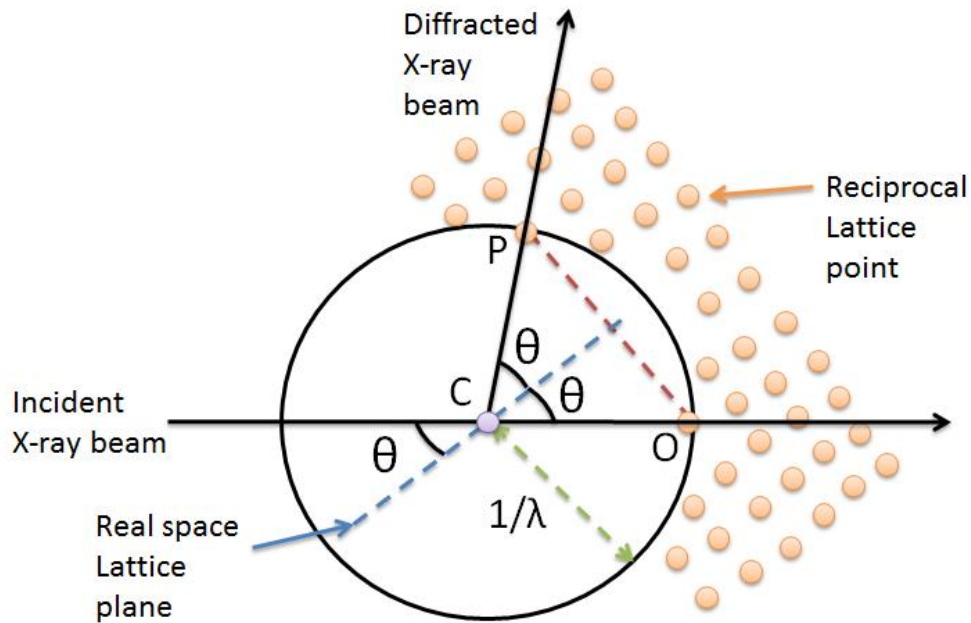


Figure 2.8: Two-dimensional slice of the Ewald sphere. A plane of reciprocal lattice (orange dots) shows a two-dimensional slice of the Ewald sphere. The spot O is the origin and reciprocal lattice points O and P are labelled as they are in contact with the surface of the sphere. A purple spot represents the centre of the sphere (C). The incident beam passes through C and exits at the origin and the diffracted beam passes through the reciprocal lattice point (P) which is in contact with the sphere with the angle 2θ . The red dotted line links OP so the blue dotted line can be constructed perpendicularly and passing through C. The radius of the Ewald sphere is the reciprocal of the incident X-ray wavelength (green dotted line).

A protein structure is a complex object and the diffraction pattern can be difficult to study, but Fourier transforms can simplify the complex wave from the diffraction pattern into amplitude and phase in the structure factor equation (Equ. 2.2). The wavelengths of the collected diffracted X-ray are the same as the incident beam but the amplitude is dependent on the density of electrons in the particular lattice plane and each diffracted beam has an independent phase angle. The quantity of f_j represents the j atom and is called the atomic scattering factor. It varies with d_{hkl} to form the Fourier transform of the electron-density distribution within the particular atom. At low resolution it equals the number of electrons in the atom, but reduces at higher resolution according to the shape of the electron distribution around the atom. The three dimensional periodic function is described with frequencies h , k and l in the x , y

z dimensions. The coordinates of the atoms in the unit cell are expressed as a fraction of the unit cell length, h_j, y_j, z_j as phase.

$$F_{(hkl)} = \sum_{j=1}^n f_j e^{2\pi i (hx_j + ky_j + lz_j)} e^{-B_j \sin^2 \theta / \lambda^2}$$

Equation 2.2: Fourier Transform linking the relationship between amplitude and phase.

The second part of the equation represents the effect of thermal and crystal lattice disorder known as the Debye-Waller function and is a Gaussian function that decreases as the angle of diffraction increase. This function would decrease as the resolution increases, proportionally to e^{-1/d^2} . The quantity B represents the breadth of smearing, and if the atom has a root-mean-square displacement u , then $B=8\pi^2(u^2)$. Both the B factor and the atomic scattering factor f vary with the resolution d to reduce the structure amplitude as d increases. At high resolution it may be necessary to consider that the smearing of the atomic electron density may not be spherical which can be done by using B which varies according to direction and is known as an anisotropic B factor.

An alternative way of calculating structure factors is to use a continuous function of electron density, $\rho(x,y,z)$ (Equ. 2.3). The factor V represents the volume of the unit cell, it arises because ρ is a density expressed as electrons per unit volume, while the structure factor F is normally expressed as electrons per unit cell. These integrations are taken over 0 to 1 in each of x, y and z , so as to extend it over the whole unit cell.

$$F_{(hkl)} = V \int_0^1 \int_0^1 \int_0^1 \rho(x, y, z) e^{2\pi i (hx + ky + lz)} dx dy dz$$

Equation 2.3: An alternative Fourier transform with continuous function of electron density.

To describe the positions of atoms within the unit cell, it is the $\rho(x,y,z)$ function that we wish to calculate. The Fourier transform of equation 2.3 allows us to determine this (Equ. 2.4) when we know the amplitudes and phases of the diffracted X-ray reflections.

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i (hx + ky + lz)}$$

Equation 2.4: Fourier transform equation for describing the position of atoms.

There are a fixed number of terms in the Fourier series which is determined by the resolution range hence the recorded reflections and so it is a summation over all combinations of h , k and l . This equation can be rearranged to separate the amplitude and phase components of the structure factors, where $|F_{hkl}|$ is the amplitude and $i\alpha_{hkl}$ is the phase contributions to the structure factor (Equ. 2.5).

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i (hx + ky + lz) + i\alpha_{hkl}}$$

Equation 2.5: Rearranged equation of Fourier transform to separate amplitude and phase.

Information can be gathered by experiment such as intensity of the reflections and the reciprocal lattice and cell volume, however, the phase parameters are usually not obtained and need to be calculated by other techniques before an electron density map can be built.

2.2.4 Data Collection and Processing

2.2.4.1 Data Collection

During experiments, radiation damage can occur, as well as loss of solvent of crystallization, and heating of the crystal which causes movement of atoms in the crystal. This damage leads to a change in Bragg's reflection intensity as a function of time, but this affect is greatly reduced by performing the experiment in nitrogen at 100 K. Thus, generally in an X-ray diffraction experiment, protein crystals are soaked in cryoprotectant solution and looped out by a small nylon loop then flash-cooled in a stream of nitrogen.

In order to measure the intensity of every unique reflection from the atoms in each plane, the crystal is rotated in the beam which places different reciprocal lattice points into reflecting positions. Furthermore, depending on the crystal symmetry, additional rotation is required to measure all the unique reflections. During each exposure the crystal is rotated around a single axis, typically by an angle between 0.1 to 2°. This allows all Bragg reflections between the chosen oscillation angles to be collected. In practice, some reflections will be fully recorded reflections on one image but others will be partially recorded reflections which are captured over two or more images.

2.2.4.2 Auto-indexing and Integration of Reflections

After the data is collected a range of programmes can be used such as MOSFLM (Leslie, 1992) and DENZO (Otwinowski and Minor, 1997) to auto-index reflections and determine unit cell dimensions and internal symmetry. A few images, normally 90° apart, are indexed and a list of space groups and the corresponding unit cell dimensions are given. A space group and cell dimensions which has the highest symmetry and the lowest penalty are chosen. The position of the direct beam is defined as the origin of the diffraction pattern and is used to measure the coordinates of X and Y of the Bragg's diffraction spots. The angle of the rotation axis (ϕ) is also noted in the experiment. By using an orientation matrix and with the fixed orthogonal axis the reciprocal lattice index (h,k,l) can be related to an experimental observation, in terms of X, Y and ϕ . The orientation matrix is improved by relating this information to the entire image frames to fit the whole dataset and by including other factors such as mosaicity and crystal orientation. The recorded reflections can then be assigned to a reciprocal lattice plane (hkl), and the intensity of each reflection can be integrated through profile fitting.

2.2.4.3 Scaling Data

Due to the symmetry of crystals, reflections of each atom are measured more than once and these equivalent reflections should have the same intensity. However, because of the small changes in beam intensity and varied crystal volume at each position the intensities are not identical. Therefore scaling is needed which introduces a scale factor to the intensities of each image to overcome the errors during data

collection to have a common scale. These symmetry-related equivalent reflections are then averaged to have the same intensity.

The CCP4 programme SCALA (CCP4, 1994) is commonly used and allow the initial process of introducing a common scale factor and process partially recorded reflections over more than one image frame. The programme also monitors the quality of the data by comparing intensities of repeated measurements or symmetry equivalent reflections. Usually these values are averaged; however those images with exceptional intensities are rejected.

The quality of the data can be quickly examined by the agreement between symmetry-related reflections within the data by the R_{sym} (Equ. 2.6) value. This statistical value is calculated for the whole data range and separated by resolution hence can be used as an indication of the resolution limit for the dataset before the experimental error becomes too extreme. In addition, the ratio of $I/\sigma I$ (where ‘I’ is the estimated mean intensity) can also suggest the resolution limit of the particular dataset as the intensity should not be twice as big as the mean intensity.

$$R_{sym} = \frac{\sum_{hkl} \sum_i (I_i(hkl) - I_{mean}(hkl))}{\sum_{hkl} \sum_i (I_i(hkl))}$$

Equation 2.6: Equation to define R_{sym} . I_i is the intensity of reflection of i with indices hkl and I_{mean} is the mean intensity of all reflections with these indices.

2.2.5 Phase Information

In the case of protein molecules, the accurate phase is acquired indirectly by phasing techniques. Three techniques, molecular replacement, anomalous dispersion and isomorphous replacement are commonly used to calculate phase. The Patterson function is the fundamental concept for these methods which use experimental data to obtain phase information.

2.2.5.1 Patterson Function

The Patterson function was described in 1934 by Lyndo Patterson (Patterson 1934); it provides an important method of working back from the diffraction pattern to the crystal structure despite the lacks of phase information. The Patterson function uses the indices h, k and l which are available experimentally as the intensities of reflections and when plotted on (u, v, w) axes form a map rather than displaying an electron density. However, peaks in the Patterson map correspond to interatomic vectors and the peak heights are proportional to the product of the scattering factors of the atomic number of the atoms involved. That is, suppose that there are atoms at A, B and C in a unit cell. The function $P(uvw)$ would show peaks at U, V, W where $O - U$ is equal to $A - B$, $O - V$ is equal to $B - C$ and $O - W$ is equal to $A - C$.

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos(2\pi(hu + kv + lw))$$

Equation 2.7: The Patterson function. The coordinates (u,v,w) define a point in a Patterson map, with phase angles equal to zero and amplitude $|F_{hkl}|^2$.

2.2.5.2 Molecular Replacement

This method takes advantage of known structures with high homology to the unsolved protein structure. The atomic position of a known structure is taken into all possible orientations and compared the structure factor amplitudes calculated from the phasing model with the collected diffraction pattern. This process involves a six-dimensional search including three rotations and three translations.

Molecular replacement uses a divide-and-conquer strategy which determines the orientation and then the translation, each involving three parameters. As the Patterson map is a vector map, it displays peaks at the position of vectors between atoms in the unit cell. These vectors are a unique signature for the protein structure and can be divided into two, intramolecular vectors (self-vectors) and intermolecular vectors (cross-vectors). Intramolecular vectors depend on the orientation of the molecule and are used in the rotation function. The non-crystallographically related molecules in the unit cell have the same intramolecular vectors. Intermolecular vectors depend on the orientation and the position of the molecule relative to the symmetry elements and are used as the translation function after the orientation is known.

Rotation Function

Despite the atoms in the molecule being shifted by crystal formation the vectors between the atoms remain unchanged. Therefore, at the rotation function, the programme is trying to overlapping the self-vectors of the search model with the self-vector of the unknown protein to find the correct orientation. The search uses a smaller radius than the phase model to avoid the effect of the cross-vector. For each orientation of Eulerian angles, α , β and γ a correlation coefficient is calculated (Rossman and Blow, 1962) and the highest volume would be most likely to be the

correct solution (Fig. 2.9). The $P(\mathbf{u})$ term is the observed Patterson function and the $P_r(\mathbf{u})_r$ component is the calculated Patterson function after the rotation function (Equ. 2.8).

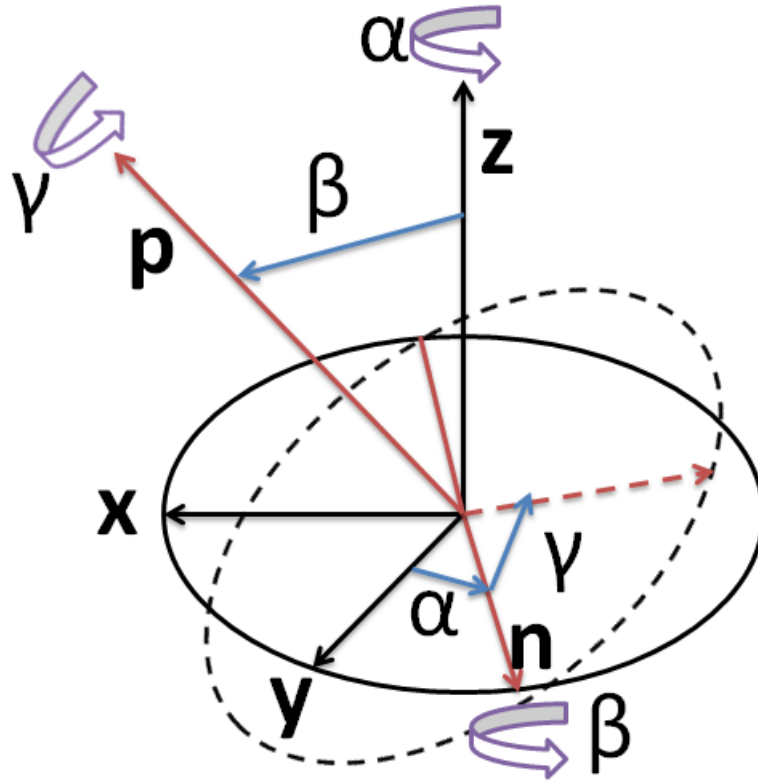


Figure 2.9: Definition of the Eulerian angles used to describe the rotation of a body. A rotation by the Eulerian angles initially rotates on the original z-axis by α , followed by a rotation on the new y-axis (n) by β then a rotation on the new z-axis (p) by γ .

$$R(\alpha, \beta, \gamma) = \int_u P(u) \times P_r(u)_r du$$

Equation 2.8: The rotation function. The fit between the Patterson density of the search model ($P(\mathbf{u})$) and the unknown structure ($P_r(\mathbf{u})_r$) is calculated with all possible rotation angles (α , β and γ). The rotation function R has a maximum value when the correct solution has been found and $\mathbf{u}$ is a point in three-dimensional space of the Patterson function.

Translation Function

After the orientation is found, the position of the molecule in the unit cell can be determined by the translation function search using cross-vectors of the Patterson function. Cross-vectors are calculated based on the chosen space group and are overlapped against the Patterson map as a function of the translation vector (t) (Crowther and Blow, 1967). As described in equation 2.9, $P_{1,2}$ are different values of the translation vectors between two symmetry related molecules and the translation function (T) will have a maximum value with the most probable solution.

$$T(t) = \int_V P_{1,2}(u,t) P(u) du$$

Equation 2.9: The translation function. The fit between Patterson density between the search model ($P(u)$) and the unknown structure ($P_{1,2}(u,t)$) is calculated and the translation function T has a maximum value when the correct solution has been found.

After the rotation and translation have been applied to the data a list of possible solutions are produced with correlation coefficient and R-factor where F_{obs} relates to the structure to be solved and F_{calc} relates to the trial model used in the searches and gives an indication of false solutions.

$$CC = \frac{\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right) \times \left(|F_{calc}|^2 - \overline{|F_{calc}|^2} \right)}{\left[\sum_{hkl} \left(|F_{obs}|^2 - \overline{|F_{obs}|^2} \right)^2 \times \sum_{hkl} \left(|F_{calc}|^2 - \overline{|F_{calc}|^2} \right)^2 \right]^{1/2}}$$

Equation 2.10: Definition of the correlation coefficient between calculated and observed structure factors for each hkl reflection. Structure factor amplitudes with a bar over them represent the average intensity for that reflection.

$$R_{work} = \frac{\sum_{hkl} \left| |F_{obs}| - k |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|}$$

Equation 2.11: Definition of R_{work} . The term k is a scaling factor that puts the calculated structure factor amplitudes on the same scale as the observed structure factor amplitudes.

2.2.5.3 Anomalous Dispersion

Another commonly used technique for phase determination is anomalous dispersion which depends on the small changes in diffraction caused by anomalously scattering atoms. As described previously, X-rays can excite atoms and cause electrons to vibrate at the same frequency causing diffraction in all directions. With this coherent scattering Friedel's law (Friedel, 1913) states that a pair of reflections h, k, l and $-h, -k, -l$ have the same structure amplitude but opposite phase (Fig. 2.10). However, when the wavelength is close to the absorption edge of an atom, its scattering does not obey this rule as the heavy atom structure factors $F_H(h, k, l)$ and $F_H(-h, -k, -l)$ have different magnitudes and phase. Absorption edge is the term describing a rapid change in the atomic absorption factor which is caused by an electron being excited to an unoccupied orbit of higher energy or ejected totally. Due to the cost in energy to excite an electron into a higher orbital, there is a reduction in the intensity of the coherent scattering. Bijvoet (Bijvoet et al, 1951) realized that when a crystal contains both normal and heavy atoms, anomalous scattering will be observed.

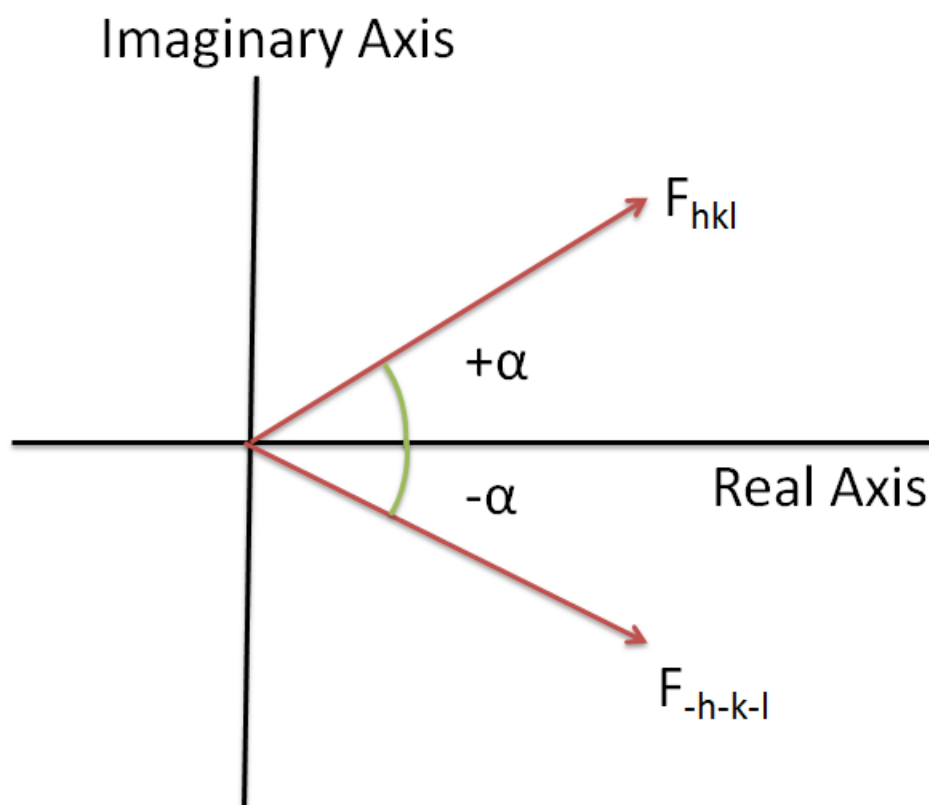


Figure 2.10: Argand diagram describing Friedel's law. Normally, the reflection of an atom will obey the Friedel's law in which the pair of reflections are related by an inversion through the origin with the same magnitude (amplitude) but opposite phases (α).

The whole atomic absorption spectrum of an element in the X-ray region includes a series of absorption edges corresponding to different shells of its electrons (*K*, *L*, *M* etc.). In a protein crystal carbon, nitrogen and oxygen are commonly found but are not near the required wavelength, but covalently bound elements such as sulphur and phosphorous are within the *K* edge range. In addition, selenomethionine and bromouracil can be used through the covalent bound method as well as ionically bound heavy metal soaks.

The normal scattering follows Friedel's law reflect on the real axis but the anomalous scattering reflect on the imaginary axis with different magnitude and phase known as the Bijvoet pair. The phase information is acquired from the differences between these Bijvoet pairs. The anomalous scattering atoms can also be located within the

unit cell and with the differences between the pairs, the Patterson map can be produced with the peaks correspond to the interatomic vectors between the pairs of anomalous scattering atoms.

With the initial positions of the heavy atoms, the occupancy and B-factors can be refined. Then the phase angle is acquired from Harker constructions (Harker 1956) and with single wavelength data set, SAD (single wavelength anomalous diffraction) is performed to obtain the phase information. With SAD experiments, two possible phase angles are calculated and depending on the signal to noise ration from the anomalous scattering atoms, more datasets might be needed to improve the phase estimates. As the results, MAD (multiple wavelength anomalous diffraction) is more commonly used which extra data sets are collected at different wavelengths hence more Harker constructions can be produced which the correct phase will have the highest score.

2.2.5.4 Isomorphous Replacement

Isomorphous replacement is based on the comparison of two crystals: a native and a derivative crystal containing a heavy metal ion or compound which binds tightly to one or multiple sites in the crystal. However, both crystals must share the same unit cell dimensions, symmetry and crystal packing. Derivative crystals are made by soaking the native crystals in solutions of heavy ions (for example Hg, Pt or Au salts) without disturbing its conformation or crystal packing. Since all atoms in the crystal will contribute to each reflection in the diffraction pattern, as long as the crystal remains isomorphous the derivate crystal can provide measureable differences in intensity compared with the native crystal and thus contribution from just the heavy atoms can be determined.

After the data is collected for both native and derivative crystals, these heavy atoms can be located by calculating the differences in amplitudes ($|F_{HP}| - |F_P|$), where F_{HP} is the scattering data by derivative crystal and F_P is from the native crystal. By using the centric reflection the amplitude contribution is placed in the Patterson function and any peaks in these Patterson maps should correspond to only the interatomic vectors between pairs of heavy atoms. After the positions of the heavy atoms are located, their occupancies and B-factors can be further refined.

Similar to anomalous dispersion technique the phase angles are acquired by Harker constructions. SIR (single isomorphous replacement) can be performed with one isomorphous derivative, but to improve the phase estimate MIR (multiple isomorphous replacements) is also commonly performed which used dataset from different isomorphous derivatives. However, different derivatives might not be exactly isomorphous and affect the accuracy of phase estimate. These errors can be reduced by combining isomorphous replacement with anomalous scattering of heavy atoms also known as SIRAS and MIRAS (single/ multiple isomorphous replacements with anomalous scattering).

2.2.6 Model Refinement

After the phase information is obtained an electron density map can be constructed. However, due to the fact that there are errors in phase calculation there will be considerable errors in the model which need correcting. By improving the model structure, phases will become more accurate since the differences between the observed and calculated structure factors will be minimised. As the phase estimates become more accurate an informative electron density map will be observed which allow further improvements of the protein structure.

2.2.6.1 Cross-validation

One of the most useful and common statistical values used in refinement is the R-factor (Equ. 2.11), but it is possible to build an incorrect model with low R-factor. Therefore, a small portion of the data (5-10%) is separated and refined independently of the working dataset (90-95%) (Brünger, 1992, Kleywegt and Brünger, 1996). Each round of refinement on the model based on the main portion of the data will reduce R_{work} but it is also compared with the small portion of the data (R_{free}) to avoid bias.

2.2.6.2 Least Squares and Maximum Likelihood

Refinement

The method of least squares is used to minimise the squared difference between the observed and calculated structure factor amplitudes. The refinement is based on the Gaussian distribution and assumes that the acquired phase is entirely correct although it is not always the case. Therefore the maximum likelihood (Pannu and Read, 1996) method is also commonly used in which observed data are considered as probabilities and in a good model these probabilities will be consistent. In the refinement the model is adjusted to maximise the probabilities of those observations occurring.

2.2.6.3 Specific Refinement Strategies

2.2.6.3.1 Solvent Flattening

This method enhances refinement by suppressing the noise by masking around the macromolecule to create a low-resolution model. All other density outside of this is set to the solvent level which can enhance the electron density map of the model.

2.2.6.3.2 Rigid Body Refinement

After the phase information is obtained, the model may not be at the optimal position in the unit cell, particularly from molecular replacement. Therefore, the whole macromolecules are considered as one rigid body for the best agreed orientation between the calculated and observed structure factor amplitude. This method is usually performed with relatively low resolution data for the reason of radius of convergence, for instance if a 6Å data are used the atoms may only move no more than 2Å.

2.2.6.3.3 Positional Refinement

After the general position and model are enhanced individual atoms can also be improved by minimising the differences between the theoretical values of bond length, bond angle, torsion angles and van der Waals contacts (Jack and Levitt, 1978). However, depending on the resolution of the data, these restraints can be stringently or minimally comprised.

2.2.6.3.4 B-Factor Refinement

The B-factor (also known as the temperature factor) is a consequence of dynamic and static disorder in crystals. This dynamic disorder is temperature-dependent vibrations of atoms within the crystal where molecules or parts of molecules in different unit cells are not occupying identical positions but have the same orientation which caused the minor differences as the static disorder. The structure factor is dependent not just on the position of the atoms, but also their vibrational parameters, so in order to obtain the best model, the B-factors must also be refined. However, at low resolution

there may only be enough information for overall B-factor refinement, whilst at higher resolution each atom can also be refined by individual B-factors.

B-factors can be refined by either an isotropic or anisotropic distribution but generally there are only enough observations to approximate isotropic B-factors, with a Gaussian distribution centred at the mean position of the atom. The width of the distribution corresponds to the mean squared deviation ($\bar{u}^2$) Å² of the atom, where $B=8\pi^2(\bar{u}^2)$ Å² hence 4 parameters can describe an atom in the model (x, y, z, and B)

Furthermore, when more observations per atom are available, an anisotropic model of B-factor can be used. In total 9 parameters can be used to describe an atom x, y, z and 6 B-factor parameters where these 6 parameters describe the orientation and elongation of an ellipsoid as written into the structure factor equation as a symmetric matrix, U_j . However, this method would easily accumulate errors hence only suggested in high resolution data and R_{free} must be carefully monitored to avoid bias: a drop of 2% is reasonable, whilst a fall of 0.5% would be doubtful.

In intermediate, TLS (translational, librational, screw motion determination) are used for B-factor refinement as it provides three parameters to described the translational, librational and screw motions of specific segments within a structure (Winn *et al.* 2001). The TLS algorithm separates the chains into segments based on comparison of the least square residual of the input structure against the TLS calculated model. This analysis is performed independently with each chain on the structure to avoid bias.

2.2.6.3.5 Manual Modelling

The accuracy of an electron density map is influenced more by phases than amplitudes but phase is calculated directly from the model so there can easily be

model bias. This is overcome by calculating a difference map where the observed structure factor amplitudes are subtracted from the calculated structure factor amplitudes. This difference map can be created by a number of ways such as $F_o - F_c$, $2F_o - F_c$, $3F_o - 2F_c$ and sigma A weighted difference map. Furthermore, waters and ligand molecules can also be built in carefully which can also improve the overall map.

2.2.6.3.6 NCS Refinement

Non-crystallographic symmetry (NCS) refinement aligns all chains with all chains in the asymmetric unit and if the alignment score is high (>80%) then local root-mean square (rmsd) can be used to identify domains and apply weights to the atoms. The domains with similar rmsd values will be tightly restrained.

2.2.6.3.7 Map Sharpening

To improve the quality of the electron density map, map sharpening can be used to increase the signal to noise ratio. Noises found in electron density map come from errors in phases and noises in experimental data. Therefore, map sharpening reduces noise amplification while increasing the signal (Altomare *et al.* 2008; Sheldrick 2008).

2.2.6.3.8 “Jelly” Body Refinement

In low resolution data, the ratio of the number of observations to the number of adjustable parameters is very small. Therefore, the chance of biasing the model is very high, but the risk can be reduced by using restraints of other known structures. The effect of ‘jelly-body’ uses similar known structure from the database and applies restraints between the rigid body and individual atoms similar to elastic network model calculation (Schroder *et al.* 2010).

2.3 Nuclear Magnetic Resonance (NMR)

NMR is amenable to the study of smaller protein structures and technique for “foldedness”, self-association (polymerisation), interactions and dynamics. It has the advantage that experiments can be performed in the solution state, which gives easy access to studying the interactions of molecules. In the case of interaction studies, by solving the protein structure or monitoring frequency changes in NMR spectra, one can explore interaction events. In this project NMR is primarily used for interaction studies and some basic theory of NMR and relevant methods are described in this sub-chapter (Montelione *et al.* 2000; Keeler 2005; Cavanagh 2007).

2.3.1 Nuclear Spin Theory

NMR involves energy transitions in the nucleus with respect to the intrinsic physical property known as the nuclear spin angular momentum. This has magnitude and direction, and has an associated nuclear spin magnetic moment which is also a vector quantity. In quantum mechanics a particle, such as an atomic nucleus, has an associated nuclear spin angular quantum number, or spin quantum number (I), which has an integer or half integer values. A number of atoms are commonly used in protein NMR, possessing an I value of $\frac{1}{2}$ (^{13}C , ^1H and ^{15}N) or 1 (^2H).

The spin of a charged particle creates a small magnetic field around itself which is directional; hence the nuclei with spin value of $\frac{1}{2}$ can be imagined as a dipole or small bar magnet (Fig. 2.11). This character causes it to have detectable properties when it is placed in an external magnetic field.

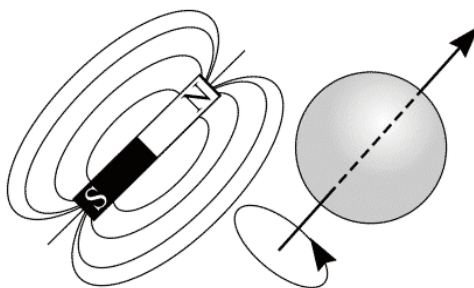


Figure 2.11: Schematic representation of the magnetic dipole of a particle. A charged nucleus possessing spin creates a magnetic moment which acts like a small bar magnet (dipole). The arrows indicate the direction of magnetic dipole.

With the magnetic properties of a nucleus, when an external magnetic field (B_0) is applied to the molecule, the nuclear dipoles of spin $\frac{1}{2}$ nuclei can align themselves either with or against the direction of the external magnetic field (Fig 2.12).

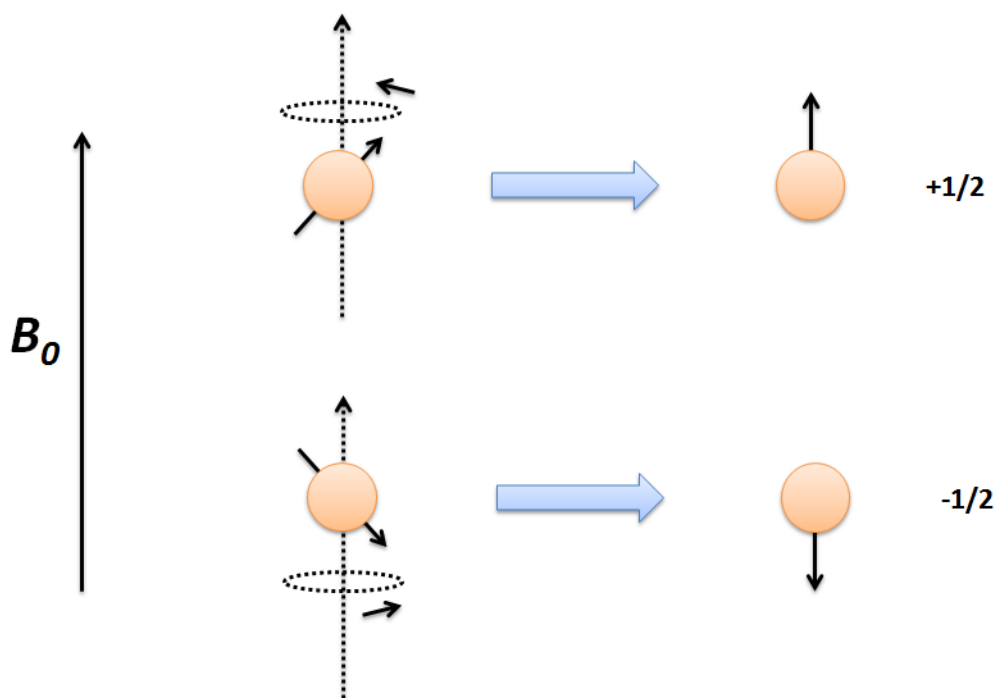


Figure 2.12: Alignment of nuclear dipoles within a magnetic field. Spin $\frac{1}{2}$ nuclei will have two orientations when an external magnetic field is applied.

2.3.2 Larmor Precession and Boltzmann Distribution

It is more energetically favourable for the dipole of the nuclei to be aligned parallel to the magnetic field. Therefore, there will be slightly more spins with this lower energy

arrangement than there are in the higher energy alignment (against the external magnetic field). This means that there is a net alignment of nuclei with dipoles aligned in the direction of the magnetic field and this ratio of the relative populations is described by the *Boltzmann distribution* (Fig 2.13).

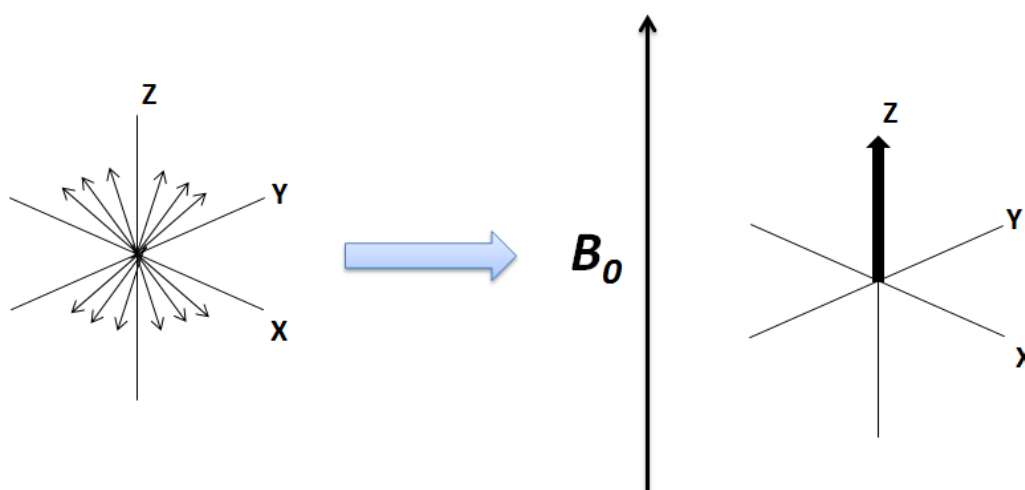


Figure 2.13: The net magnetisation of nuclei in the magnetic field. The net magnetisation of nuclei is in the same direction as B_0 .

In fact, the dipoles, rather than aligning along the axis of the external magnetic field, position themselves at an angle away from it. This effect causes them to precess around B_0 as they try to align with it. The frequency at which they precess around this axis is known as the *Larmor Frequency* (ω). This frequency can vary, depending on the type of nucleus and the strength of the external magnetic field in which it is placed. However, the net magnetisation is aligned with B_0 .

It is possible to flip the nuclear dipole in the magnetic field between the lower and higher energy state. This requires absorption of a photon of energy, where the energy required must be equivalent to the energy difference between the two states (Fig. 2.14). This corresponds to the radio frequency (RF) region of the electromagnetic spectrum.

If this radiation has the same frequency as the Larmor frequency of the nucleus, it can absorb a photon of energy and flip the dipole to the higher energy state, which can lead towards a *saturation* state. At this state there is no net magnetisation, because the number of high energy nuclei is equal to the number of low energy nuclei. When the radiation is removed, the magnetic dipoles begin to relax back to their initial population distribution due to the *spin-lattice relaxation* effect.

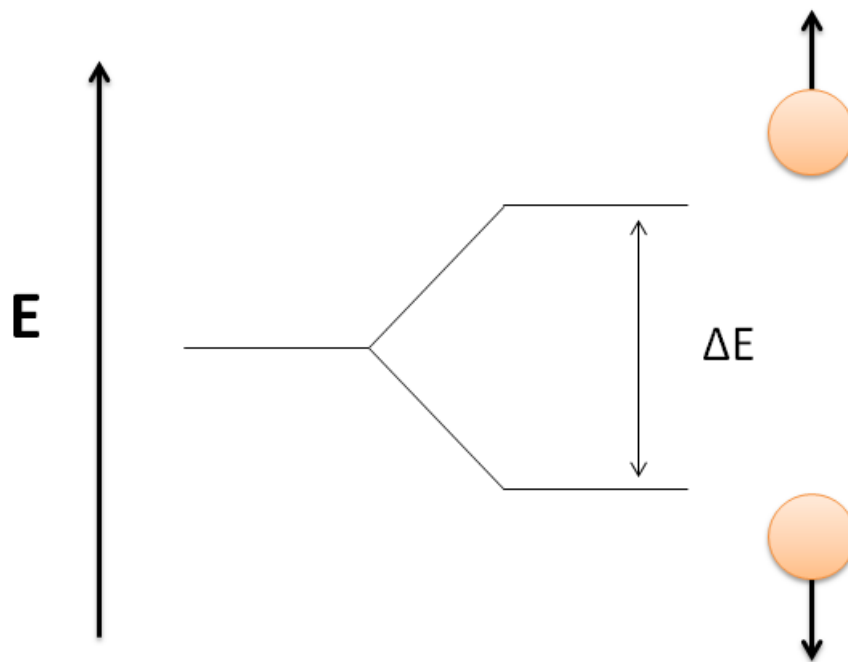


Figure 2.14: Difference in energy between the two spin states. The nuclei can absorb radiation with the correct energy ΔE , as a result changing the spin states.

2.3.3 Pulse NMR

An external pulse can move a magnetisation vector from its equilibrium position. The bulk net magnetic vector (M_0) of the nuclei is along the z axis before the pulse is applied and when the external pulse is applied the magnetisation rotates inward the x-y plane following the right-hand grip rule (Fig 2.15). The angle θ through which the vector turns is known as the tip angle and depends on the amplitude and duration of the pulse (B_1). The angle rotated by the magnetic vector is 90° if the pulse is switched off when it is in the transverse plane and magnetisation then precesses in the x-y plane around the z axis (B_0 field). This transverse magnetisation produces a non-zero magnetic moment in the x-y plane, which induces a time dependent current in the receiver coil known as the free induction decay (FID).

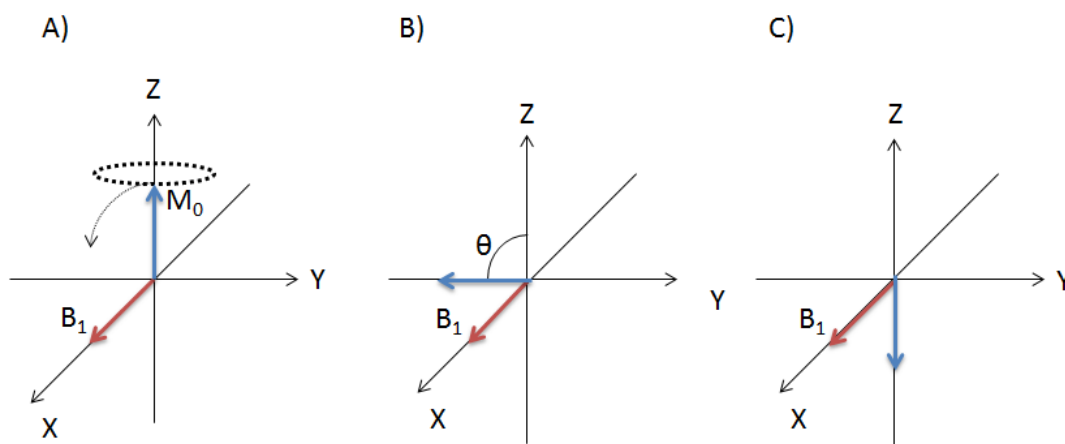


Figure 2.15: Effect of a pulse of RF on bulk magnetisation. A) B_1 pulse pushes the M_0 toward the – Y axis. B) and C) The flip can be at any angle depending on how long the pulse is applied on the X axis.

The FID from the NMR experiment contains information in the form of frequencies in the time domain and they are difficult to interpret. However, it is much easier to understand these data if they are presented in the frequency domain, which shows peaks at the NMR frequencies of the nuclei being detected. Therefore, Fourier transformation is applied to the FID and the spectrum can be displayed as a series of Lorentzian-shaped peaks (Fig. 2.16).

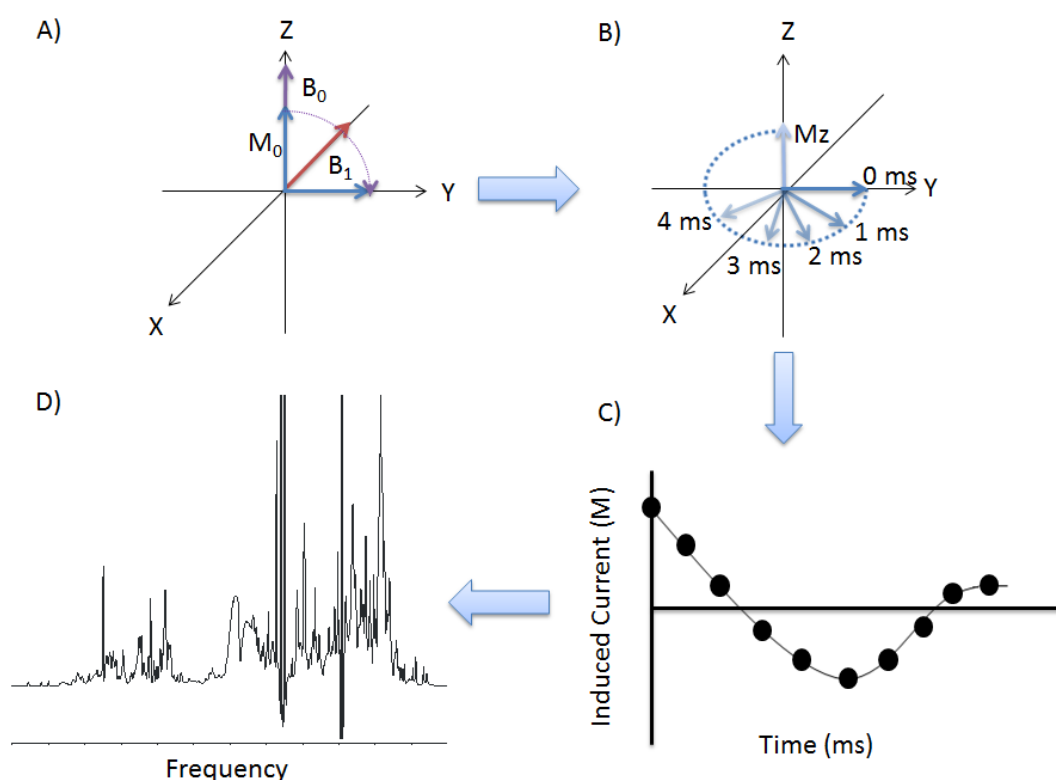


Figure 2.16: A vector model representation of the pulse experiment. A) The equilibrium bulk magnetisation (purple) is brought to the Y axis by a 90° pulse along the X axis. B) After the 90° pulse, the transverse magnetisation (blue) decays back to the initial state by precessing around the Z axis and generates a signal in the receiver coil. C) The FID is observed by detection at the Y axis and after Fourier Transform a spectrum with Lorentzian-shaped peaks is obtained (D).

2.3.4 Coupling Effects

The magnetic field of a nucleus can influence its neighbours and this can be directly through space (dipolar coupling) or through intervening covalent bonds (scalar coupling)

2.3.4.1 Scalar Coupling

Indirect scalar coupling (J coupling) can occur between two nuclei through intervening chemical bonds and the effect is normally vanishingly small above three covalent bonds. It causes a splitting effect on the neighbouring signals and can give hints on the connectivity of atoms in the protein. If a proton has a single spin-coupled neighbour, it will appear as a doublet or if there are two identical atoms then a triplet would be observed (Fig. 2.17). The distance between the multiplet components is given by the coupling constant J (Hz), which is independent of the magnetic field strength. Normally, couplings over one ($1J$) two ($2J$, germinal coupling) and three bonds ($3J$, vicinal coupling) are observed. Vicinal couplings depend on the torsion angle between the coupling protons.

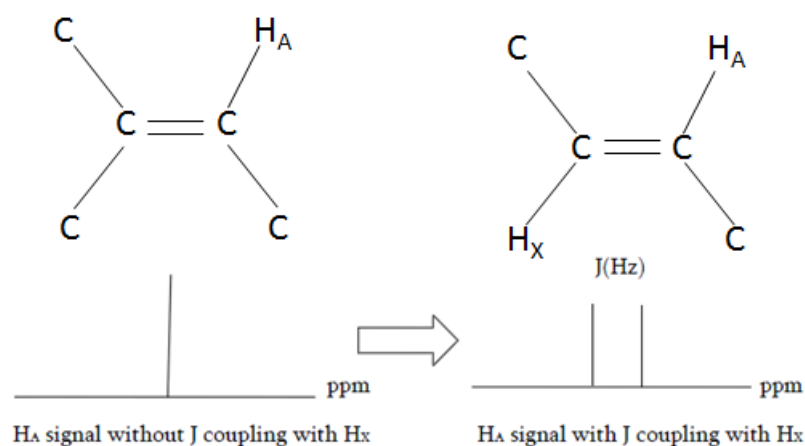


Figure 2.17: An example of J coupling effect between H_A and H_X in 2,3-dimethylbutane.

2.3.4.2 Dipolar Coupling

Dipolar coupling can occur between two spin $\frac{1}{2}$ nuclei and its effect depends on their orientation and distance in space. For instance, if nucleus A is positioned within the magnetic field of nucleus B, this can affect the surrounding total magnetic field experienced by nucleus A depending on the spin state of B. Two protons which are separated by 3 Å in space can have dipolar coupling around 4.5 kHz, but this effect decreases rapidly with an increase of distance between dipolar coupled nuclei. In solution, the rapid tumbling of the protein molecules averages the overall dipolar coupling to zero. However, if anisotropic media is used to achieve a partial “alignment” of the protein, a small proportion of the dipolar coupling can be regained, called the residual dipolar coupling (RDC). This gives information about the angle of the internuclear vector relative to the external magnetic field, hence the relative orientation of parts of the molecule (Clore and Gronenborn 1998).

In a folded protein, one nucleus would be affected by its near non-equivalent nucleus. The transfer of spin polarisation from one spin population to another via cross-relaxation is known as the Overhauser Effect and is affected through space rather than chemical bonds. This effect is commonly used and measured by the technique called NOESY (Nuclear Overhauser Effect Spectroscopy) which gives information about local structure or intermolecular interactions and assists in structure determination (Neuhaus D 1989).

2.3.5 Chemical Shift

In a molecule each atomic nucleus can have a unique pattern of energy levels and resonance frequencies. This is because, even for the same type of nucleus, the electron distribution can vary depending on the local environment due to the electronegativity

of neighbours, bond lengths and torsion angles. The variation in their unique nuclear magnetic resonance frequencies of the same nucleus is conventionally described by the chemical shift. The chemical shift is a magnetic field strength independent measure of the frequency of a nucleus due to the influence of its local chemical environment and is measured in parts per million (ppm) relative to a reference frequency. The shift can be affected through effects occurring both intra-molecularly and inter-molecularly and thus can give an indication of intermolecular interactions, e.g. a ligand binding to a protein.

2.3.5.1 Electron Shielding

The external B_0 field induces an electric current in the electron distribution around a nucleus, which generates a magnetic field that opposes B_0 . Consequently the field felt by the nucleus is changed due to this shielding effect, and hence its Larmor frequency, which is proportional to field strength. The electron distribution around a nucleus is influenced by the electronegativity of covalently bound atoms, which in turn influence the amount of electron shielding experienced by the nucleus and gives some clues as to the chemical (covalent) environment of a nucleus. An example is the chemical shift difference between amide and methyl protons. As nitrogen is more electronegative than carbon, amide protons are less shielded than methyl protons and hence the amide protons have higher Larmor frequencies. The chemical shift of amides is observed around 7 ppm whereas methyl protons are observed around 1 ppm.

2.3.5.2 Ring Current Shifts

The delocalised electrons of aromatic rings are a major source of local magnetic fields which generate a magnetic field that strengthens B_0 in the same plane as the aromatic ring but opposes in planes above and below. In proteins the ring protons of the

aromatic residues thus experience a reinforcement of B_0 and resonate at around 7 ppm. In folded proteins aromatic residues are likely placed in a stable hydrophobic core packed against aliphatic residues hence the methyl groups of aliphatic residues can resonate around 0 ppm or -1 ppm rather than the usual 1 ppm. This is a good indication of the folded state of a protein.

2.3.6 Relaxation

After the excitation pulse, the system needs to re-establish its equilibrium, and this process is known as relaxation. There are two different types of relaxation events, longitudinal relaxation (T_1) and transverse relaxation (T_2).

Longitudinal relaxation, also known as Spin-Lattice relaxation, occurs after pulse excitation when the bulk magnetisation is returning to its net alignment along the z axis due to the re-establishment of the initial state population (Fig. 2.18). During longitudinal relaxation, the energy is transferred to the surroundings as heat but is too small to be detected. The rate of this process (R_1) is usually milli-seconds to seconds in protein solution NMR but is affected by a number of factors including molecular size, temperature, solution viscosity and molecular shape.

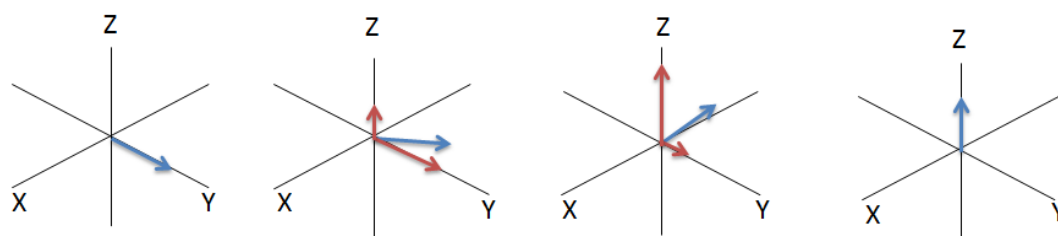


Figure 2.18: Longitudinal Relaxation Process. The bulk magnetisation (blue) is returning to Z axis as the excited magnetisation decreases on Y axis and increase on Z axis (red).

In addition, the re-equilibration process also involves another event, transverse relaxation, also known as Spin-Spin relaxation. Transverse relaxation is the loss of phase coherence of spins in the x y plane and is caused by the slight differences in the magnetic field strength experienced by a spin over the sample because of the orientation dependence on the local field (Fig. 2.19).

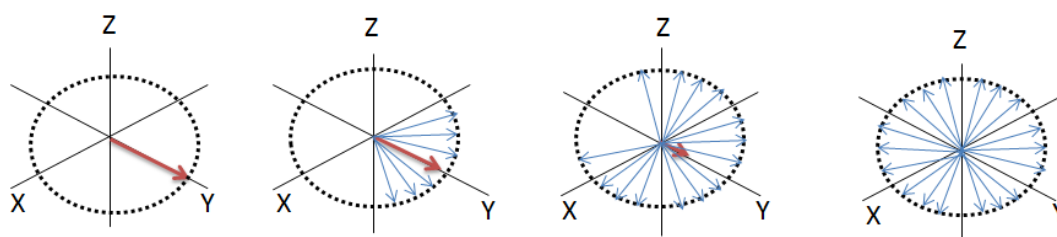


Figure 2.19: Transverse Relaxation Process. The bulk magnetisation is coloured red and the individual spins are coloured in blue. The equilibrium bulk magnetisation is decreasing while the individual magnetisations re-establish equilibrium.

2.3.6 One Dimensional NMR

Each 1D NMR experiment involves two stages: preparation and detection. During the preparation stage the spin system is allowed to reach its equilibrium and then a RF pulse is induced for sufficient time and power to rotate the magnetic vector from the z axis to the x-y plane. After the pulse, the precessing vector returns to equilibrium along the Z axis inducing a fluctuating and decay in voltage detected in the receiver coil. The same procedure is usually repeated to cancel out randomly occurring noise using signal averaging to achieve a better signal to noise ratio. A good signal to noise ratio would suggest an appropriate protein concentration (typically 0.5 mM) which would become more important in further experiments.

1D NMR can give some indication of the state of protein despite the fact the signals are extensively overlapped. As the result of chemical shift, the positions of peaks from the methyl and amide groups are a good indication of folded protein. In addition,

resonance line width gives indication of any self-association or multimerisation as the larger protein size increases tumbling time which shortens the T_2 relaxation time.

2.3.7 Two Dimensional NMR

A basic 2D experiment consists of 4 stages: preparation, evolution, mixing and detection (Fig. 2.20). During the preparatory period the samples are excited by one or more pulses and the magnetisation is evolved in the evolution period. This evolution period is also known as the t_1 period and the duration increases in the course of the 2D experiment to encode the frequency of the nucleus excited during the preparation phase. The mixing period comprises further pulses where magnetisation is transferred to another nucleus and the FID is recorded in the detection period also known as the t_2 period. With Fourier transform applied to both t_1 and t_2 , to f_1 and f_2 respectively, a 2D plot can be produced of the frequency domain signals and the intensities of the signals in the third dimension.

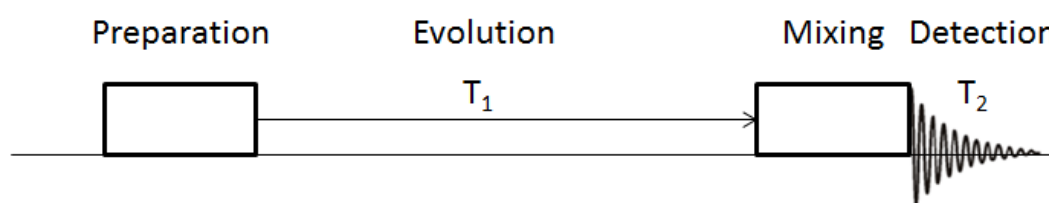


Figure 2.20: A schematic for 2D NMR Spectroscopy.

2D experiments can be used to correlate ^1H with an attached heteronucleus (e.g. ^{15}N , ^{13}C). In proteins, this is conventionally achieved by using the isotope ^{15}N to perform ^{15}N -HSQC. In the 2D ^1H ^{15}N -HSQC experiment, the amide group of all amino acids (except proline) are represented by a peak and additionally from the side chains of arginine (one peak), tryptophan (one peak), glutamine (two peaks) and asparagine (two peaks). The 2D ^1H ^{15}N -HSQC experiment is useful for monitoring the state of a

protein as any change in the chemical environment such as binding to a ligand can be detected rapidly at the residue-level (Keeler 2005).

2.3.8 Three Dimensional NMR

Monitoring the protein in 2D spectra is very useful for interaction studies but each peak also needs to be “assigned” in order to determine the frequency of each atom. 3D experiments are commonly needed and conventionally the protein samples are produced with ^{13}C and ^{15}N isotopes for backbone assignment (Gardner and Kay 1998). Various experiments can be performed but typically HNCACB, CBCA(CO)NH, HN(CA)CO and HNCO are used to determine chemical shifts of atoms in the backbone of a protein. Other side chain experiments are not discussed here because in this project NMR is used for interaction studies only requiring backbone assignment.

The HNCACB correlates the amide protons and nitrogen with the additional third dimension of the spectrum containing the ^{13}C chemical shift of the C^α and C^β resonances of a given residue and of the residue before it in the sequence (Fig. 2.21).

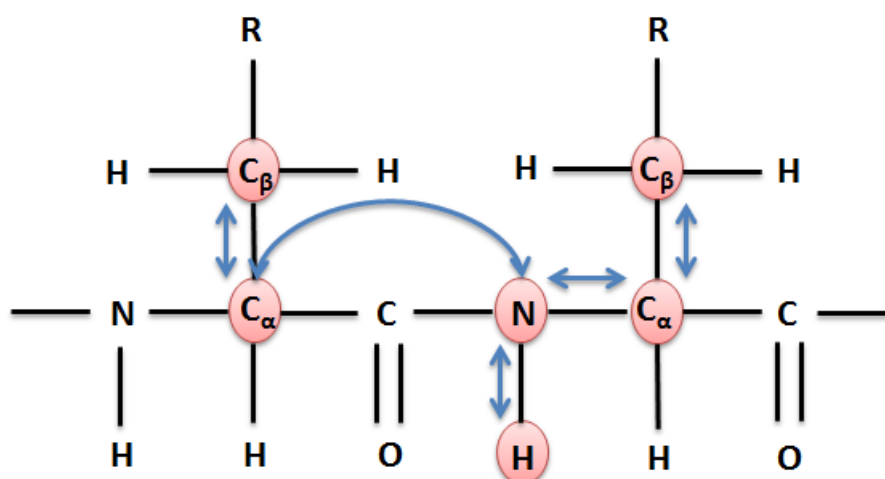


Figure 2.21: The magnetisation transfer in the HNCACB experiment. Blue arrows indicate the pathway of magnetisation transfer.

In CBCA(CO)NH experiments, magnetisation is transferred first from the H^α and H^β to the protons C^α and C^β via J-coupling then, further transfer via the carbonyl group (CO) to the NH in the amide group (Fig. 2.22). The CBCA(CO)NH thus displays peaks (C^α and C^β) only from i-1 as the direction of magnetisation indicates. It involves fewer transfer steps than the HNCACB hence it is more sensitive.

In a given strip of HNCACB, it shows the C^α and C^β from a particular residue (i) and proceeding residue (i-1) and by comparing with the strips from CBCA(CO)NH the backbone of the protein can be assigned.

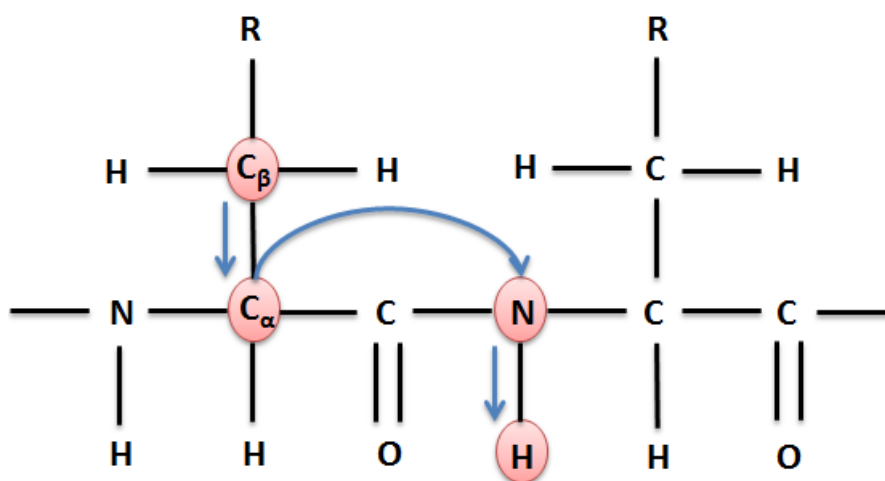


Figure 2.22: The magnetisation transfer in the CBCA (CO) NH experiment. Blue arrows indicate the pathway of magnetisation transfer.

The HN(CA)CO experiment is specifically designed to correlate inter- and intra-residue backbone connectivity between the amide ^{15}N - ^1H pair and the carbonyl ^{13}CO resonance. The magnetisation is transferred from the proton in the amide group to C_α of residue i and i-1 via the J-coupling before transfer to CO for frequency encoding. The signal is then transferred back via the same pathway to record the N frequency before signal detection on the NH proton (Fig. 2.23).

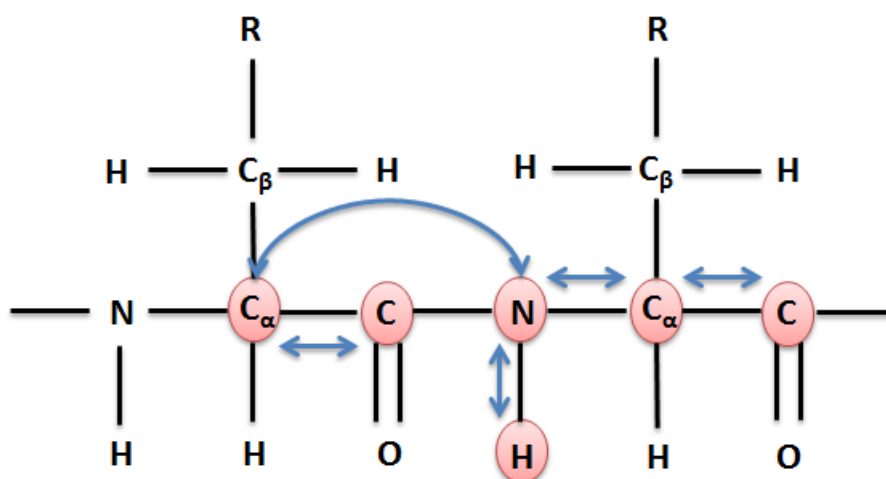


Figure 2.23: The magnetisation transfer in the HN (CA) CO experiment. Blue arrows indicate the pathway of magnetisation transfer.

In conjunction with HN(CA)CO, the HNCO is often acquired. It gives information on the backbone connectivity between the amide group with the carbonyl resonance of the preceding residue, allowing CO peaks from residue *i* and *i*-1 in the HN(CA)CO to be distinguished (Fig. 2.24).

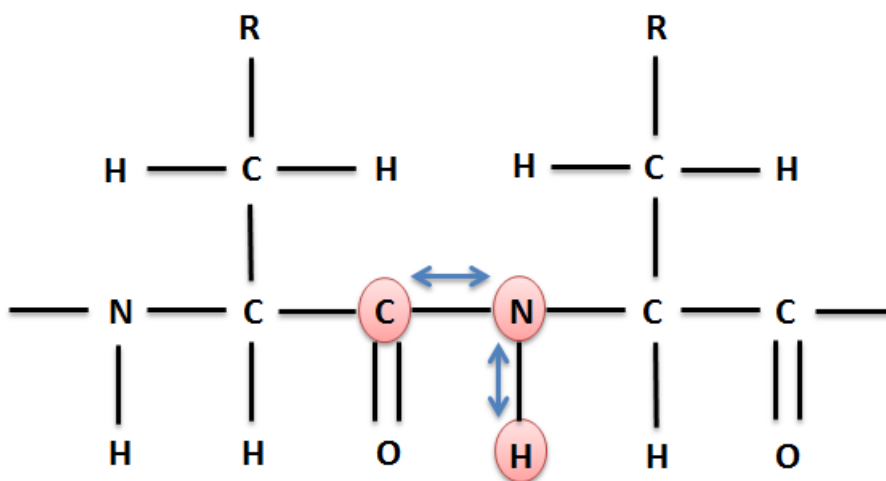


Figure 2.24: The magnetisation transfer in the HNCO. Blue arrows indicate the pathway of magnetisation transfer.

2.4 Summary

X-ray crystallography is commonly used to determine high resolution structures and complement to NMR spectroscopy for structural and functional studies. NMR spectroscopy is a versatile technique for studying protein-ligand interactions. For more detailed analysis, resonance assignment is required and this is conventionally achieved with ^{15}N and ^{13}C isotope labelled protein.

In this project, the structure of AafB and HdaB was successfully determined by X-ray crystallography using the molecular replacement phasing technique (Chapter 5 and 7). AafB was not yet co-crystallised with its potential ligands. Therefore, the complementary technique of NMR was used for interaction studies with the potential oligosaccharides (Chapter 6).

Chapter Three: Materials and Methods

3.1 Recombinant DNA Techniques

3.1.1 Materials (DNA templates)

The DNA sequences were obtained from the GenBank corresponding to the desired protein sequences and were improved by optimization of codon usage for expression in *E. coli*. The donor strand complemented (DSC) constructs were designed based on the sequenced alignment within the family, and the N-terminal donor β -strand was removed and placed on to the C-terminus of each protein and were sent for synthesis by Mr. Gene then inserted into a PQE-30 plasmid (see appendix 1 for full sequence alignment).

- *aafB-dsc-afaE* – A native *aafB* sequence was linked with the donor β -strand from *afaE* on the C-terminus.
- *afaDE-VII* – A dimer of *afaD-VII* and *afaE-VII* was designed to have a donor β -strand from *afaE-VII* on both C-termini and connected by a small peptide.
- *aggBA3* – Both native *aggB3* and *aggA3* were inserted with a donor β -strand from *aggA3* on their C-termini and joined as a dimer.
- *hdaBA* – A donor β -strand of *hdaA* were placed on the C-termini of both proteins and then joined by a linker to form a dimer.
- *sefDA* – A linker was designed to join the two native sequence of *sefD* and *sefA*, in which both were engineered to have a donor β -strand from *sefA* on their C-termini.

The *aafB-dsc-aafA* was cloned from *aafB-dsc-afaE* and the donor β -strand was replaced from *afaE* to *aafA*.

The donor strand complemented version of *aafA* was provided by Dr. Yi Yang (Imperial College) and was further engineered to make a dimer of *aafB* and *aafA* by introducing a linker in between as named, *aafBA*.

3.1.2 Primers Design

All primers were synthesised by MWG (Eurofins), the corresponding product sizes and sequences are listed in appendices 2. All primers were designed to include the appropriate restriction sites on the 5' end and the 3' end with low GC contents and stop codon (TAA) where is required (Appendix. 2).

3.1.3 Polymerase Chain Reaction (PCR)

Reactions were made up as described in Table 3.1 and cycling was run as follows:

1. 94 °C for 1 min
2. 94 °C for 30 sec
3. 50 °C for 30 sec
4. 72 °C for 1 min
5. go to step 2 and repeat for 25 cycles
6. 72 °C for 5 min

Total Volume	100 μl
<i>Pfu</i> Buffer (10x)	10 μ l
dNTPs (2.5 mM)	10 μ l
DNA template (20 ng/μl)	2 μ l
Forward oligo (5 ng/μl)	4 μ l
Reverse oligo (5 ng/μl)	4 μ l
ddH₂O	69 μ l
<i>Pfu</i> DNA polymerase (Promega)	1 μ l

Table 3.1: PCR Reaction recipes

3.1.4 DNA Purification

3.1.4.1 PCR Purification

PCR purification was performed with a kit from Qiagen, which sufficiently removed all unwanted buffers and impurities.

3.1.4.2 DNA Agarose Gels

1.2% (w/v) agarose gels were made up with TAE buffer (40 mM Tris; 0.114% glacial acetic acid; 1 mM EDTA, pH 8.0) and SYBR safe (Invitrogen). Samples were mixed with 6x loading dye (2.5% ficoll 400, 11 mM EDTA, 3.3 mM Tris-HCl, 0.017% SDS and 0.015% bromophenol blue at pH 8.0) and run at 100 V for 30 minutes. The DNA bands were visible under blue-light transilluminator. 1 Kb and 100 bp DNA ladders were obtained from New England Biolabs.

3.1.4.3 DNA Extraction from Agarose Gels

The DNA bands with the appropriate size were carefully cut out and the DNA was extracted using and following the protocol of the QIAquick Gel Extraction Kit from Qiagen.

3.1.5 Restriction Endonuclease Digestions

Both restriction endonuclease enzymes and the appropriate buffers were purchased from New England Biolabs. Double digestion with two restriction endonucleases was performed with the appropriate buffer and the samples were left at 37 °C for 2 hours (Table 3.2).

Restriction Enzymes	BamHI and HindIII	AleI and HindIII	AleI and BamHI
Buffer	2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl ₂ , 1 mM Dithiothreitol at pH 7.9)	4 (50 mM Potassium Acetate, 20 mM Tris-Acetate, 10 mM Magnesium Acetate, 1 mM Dithiothreitol at pH 7.9)	4

Table 3.2: Digestion buffer used in double digestion with different restriction enzymes.

3.1.6 DNA Ligation

All double digested DNA products were ligated to PQE-30 plasmid which has a T5 promoter for high level of expression and a carbenicillin resistance gene for positive

screening (Fig. 3.1). T4 DNA ligase (New England Biolabs) and T4 ligase buffer were also added and the samples were left at room temperature for 5 minutes.

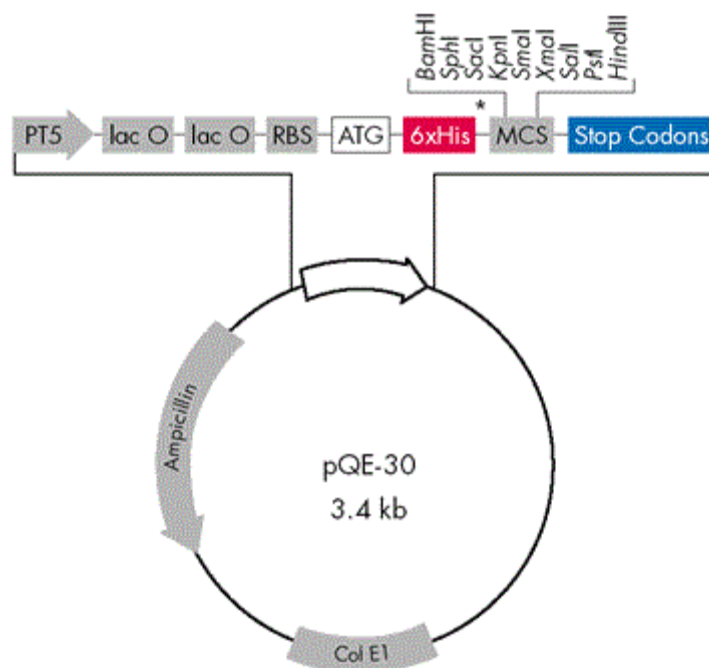


Figure 3.1: The plasmid map pQE-30.

3.1.7 Transformation of *E. coli* Strains

1-5 μ l DNA were mixed with 50 μ l XL1 Blue supercompetent cells (Qiagen), XL1 Blue competent cells (Qiagen) or *E. coli* strain M15 (with pREP4 plasmid for expression where pREP4 could efficiently repress protein expression prior to IPTG (isopropyl-beta-D-1-thiogalactopyranoside) induction and selected by kanamycin) and left on ice for 30 minutes. Heat shock was then applied to the sample at 42 °C for 45 seconds and left on ice for an addition 5 minutes. Samples were spread onto LB plates (10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl and 15 g/L Agar) with the appropriate antibiotics overnight at 37 °C.

3.1.8 Plasmid Preparation

Plasmids were purified by using and following protocol of QIAprep Spin Miniprep Kit from Qiagen.

3.1.9 Sequence Analysis

DNA was sent to a commercial sequencing centre, Beckman Coulter Genomics for sequencing with the universal forward and reverse primers of T5 promoter. The sequencing results were aligned against the original gene sequence of interest and confirmed of no mutation event.

3.2 Protein Techniques

3.2.1 Protein Expression

A colony was picked and allowed to grow in 50ml LB at 37 °C overnight with the appropriate antibiotics as a start-up culture. A pellet of cells was harvested from the start-up culture by centrifuging at 2000 rpm for 10 minutes at room temperature. The pellet was gently resuspended and introduced into fresh media with the appropriate antibiotics. Cultures were grown in 1 L LB with the appropriate antibiotics at 37 °C in an orbital shaker to OD_{600nm} 0.4-0.6 and then IPTG was added to a final concentration of 500 µM. The culture was then allowed to grow at 37 °C in an orbital shaker for 4 hours before harvesting the cells.

In addition, M9 minimal media was used to produce a ¹⁵N or ¹⁵N, ¹³C labelled protein samples where ¹⁴N ammonium chloride and ¹²C *D*-glucose were substituted by ¹⁵N ammonium chloride and ¹³C *D*-glucose respectively (Spectra Gases).

3.2.2 Protein Purification

One litre of culture was centrifuged for 10 minutes at 4,000 rpm and the pellet was resuspended in 25 ml of denaturing binding buffer (50 mM NaPO₄, 300 mM NaCl, 8 M urea, 5 mM imidazole pH 8.0) and one EDTA free protease inhibitor cocktail tablet (Roche). The cells were lysed by French Press (SLM Instruments) and centrifuged at 16,000 rpm (SS-34 rotor, Sorvall) for 30 minutes at 4 °C. Ni-NTA agarose resin (Qiagen) was equilibrated with 20 ml of binding buffer before applying the supernatant. The resin was washed with 50 ml of washing buffer (50 mM NaPO₄, 300 mM NaCl, 8 M urea, 10 mM imidazole pH 8.0) and then the protein was eluted with 20 ml elution buffer (50 mM NaPO₄, 300 mM NaCl, 8 M urea, 50 mM imidazole pH 8.0).

3.2.3 Protein Refolding

The purified protein was refolded by dialysis using dialysis membrane with 3500 Da MWCO (Spectrum) against four litres of refolding buffer (50 mM NaCH₃CO₂, 200 mM NaCl pH 5.0, 1 M urea) at 4 °C overnight with mild stirring. Further dialysis was carried out with the same buffer but without any urea (50 mM NaCH₃CO₂, 200 mM NaCl pH 5.0) at 4 °C overnight with gentle stirring to complete refolding the protein sample. Protein samples were further concentrated by Vivaspin 20 centrifugal concentrator with 3000 Da MWCO (Sartorius) to 1 ml.

3.2.4 Protein Chromatography

Gel filtration was performed on a SuperdexTM75 prep grade column (GE Healthcare) using an AKTATM FPLC (Amersham Biosciences) to enhance the purity of the protein samples. Protein samples were run in 50 mM NaCH₃CO₂, 200 mM NaCl pH

5.0 and the molecular weight was determined by protein marker (Sigma). The final protein samples were collected and concentrated by Vivaspin 20 centrifugal concentrator with 3000 Da MWCO (Sartorius).

3.2.5 SDS-PAGE

Protein samples were mixed with Novex Tricine SDS sample buffer (Invitrogen) in the ratio of 1:1 and then heated up to 100 °C for 5 minutes before loading on to a 15% acrylamide tricine gel in a BioRad vertical electrophoresis unit to run for 30 minutes at 180 V. The 15% acrylamide tricine gels were prepared using the recipe developed by Schagger *et al* 1987 (Schagger and von Jagow 1987). The gels were either stained in Coomassie blue or InstantBlue (Expedeon) and the molecular weight of each protein band was determined by a protein maker (Invitrogen).

3.2.6 UV Spectroscopy

The concentration of each protein sample was quantified by UV spectroscopy. Nanodrop (Thermo Scientific) was used to detect the absorbance of the protein samples at the wavelength of 280 nm and the concentration calculated using the Beer's law (Equ. 3.1). A is the absorbance at 280 nm, ϵ is the extinction coefficient, c is the protein concentration in molar and l is the length of the light path. The extinction coefficient of all protein samples were estimated by ProtParam from ExPASy.

$$A = \epsilon \cdot c \cdot l$$

Equation 3.1: Beer's law.

3.3 Nuclear Magnetic Resonance Experiments

3.3.1 Sample Buffer

All samples for NMR were buffer exchanged using a Vivaspin 20 centrifugal concentrator with 3000 Da MWCO (Sartorius) into NMR buffer (50 mM NaCH₃CO₂, 50 mM NaCl, 10% D₂O pH 5.0)

3.3.2 NMR Data Acquisition

Data were acquired on Bruker DRX 500 and 600 MHz Spectrometers with TXI cryoprobes (Imperial College London) at 298 K. 1D NMR spectra were processed using Topspin software (Bruker Biospin Ltd) and ¹H-¹⁵N HSQC, titration NMR and 3D data were processed using NMRPipe and analysed in NMRView 5.0 (One Moon Scientific, INC) (Delaglio *et al.* 1995). The acquisition parameters of each experiment are listed in Table 3.3. All one-dimensional ¹H NMR experiments were recorded using a Hahn-Echo with water pre-saturation or excitation sculpting with gradient water suppression (Hwang 1995).

The backbone assignment was initially carried out by MARS, automatic assignment software followed by several rounds of manual assignment (Hwang 1995; Jung and Zweckstetter 2004).

Experiment	Nucleus	Spectral Width /Hz	Temperature (K)
HSQC	^1H	6250	298-313
	^{15}N	1666.67	
HNCACB	^1H	6250	298
	^{15}N	1666.67	
	^{13}C	8333.33	
CBCA(CO)NH	^1H	6250	298
	^{15}N	1666.67	
	^{13}C	8333.33	
HNCO	^1H	6250	298
	^{15}N	1666.67	
	^{13}C	8333.33	
HNCA(CO)	^1H	6250	298
	^{15}N	1666.67	
	^{13}C	8333.33	

Table 3.3: NMR experiments and acquisition parameters.

3.4 X-ray Crystallography

3.4.1 Pre-Crystallization Tests

Pre-crystallization tests (Hampton Research) were performed with all the proteins in acetate buffer (50 mM NaCH₃CO₂, pH 5, 50 mM NaCl) to estimate the appropriate protein concentration for crystallization trials.

3.4.2 Crystallization Screens

Crystallization conditions were screened using 200 nl drops (1:1 protein:well solution) on a Mosquito crystallization robot (TTP labtech) with commercial reagents from Hampton Research (crystal screen 1 & 2, PEG/Ion & Natrix, Index, SaltRX and PEG/Ion2), Emerald BioSystem (Wizard Screen 1, 2 and 3) and Molecular Dimensions Ltd (Mem-Start & Mem-Sys, PACT, JCSG+ and MemGold). IZIT crystal dye (Hampton Research) was used to differentiate protein from salt crystals. Further optimisation of these conditions was performed by varying pH and the concentration of precipitants. The samples were allowed to crystallize by using either the sitting drop or hanging drop method of vapour diffusion with 1 ml reservoir and 6 µl drops (1:1 protein:well solution).

3.4.3 Cryoprotection

Protein crystals were directly transferred from optimization trays to the appropriate cryoprotection solution which consisted of the well solution and cryoprotection compounds (Table 3.4) depending on the crystals before placing them under a stream of nitrogen at 100 K. In addition, the appropriate amount of sodium iodide and 5-amino-2, 4,6-triiodoisophthalic acid (I3C) was also added to the cryoprotectant when needed.

Protein	Well Solution	Cryoprotectant
AafB	pH 4.9-pH 6.3 of 0.1 M BIS-TRIS, 0.2 M Lithium sulphate, 23%-28% Polyethylene glycol (PEG) 3350	pH 5.3 of BIS-TRIS, 0.2 M Lithium sulphate, 40% Polyethylene glycol (PEG) 3350, 20% Polyethylene glycol (PEG) 400
HdaB	pH 4.5-pH 5.9 of 0.2 M Ammonium Citrate, 18%- 23% Polyethylene glycol (PEG) 3350	pH 4.8 of Ammonium Citrate, 40% Polyethylene glycol (PEG) 3350, 20% Polyethylene glycol (PEG) 400
SefD	pH 4.0-pH 5.4 of 0.1M Sodium Acetate, 3.7 M-4.2 M Ammonium Acetate	pH 4.0 of 0.1 M Sodium Acetate, 3 M Ammonium Acetate, 30% Glycerol

Table 3.4: Composition of different cryoprotection solutions. Different solutions for different protein crystals, which consists of well solution and cryoprotectant.

Chapter Four: Cloning, Expression, Purification and Crystallisation

4.1 Cloning

Based on the sequence alignment (Appendix 1) of AfaD homologues, a phylogram (Fig. 4.1) was constructed where the similarity level varied from 23% to over 70%. The phylogram shows various branches which suggested these homologous proteins are fairly diverse. We have chosen a protein from each branch to cover the diversity of sequence. These proteins, Afa-VII, Sef, Agg3 and Hda have similarities ranging from 23% to 70% and all are encoded on a similar chaperone-ushe operon and are involved in fimbrial formation (Clouthier *et al.* 1994; Lalioui *et al.* 1999; Boisen *et al.* 2008; Touchon *et al.* 2009). The DNA sequence of each protein was acquired from the GenBank and was further optimized for codon usage to improve protein expression.

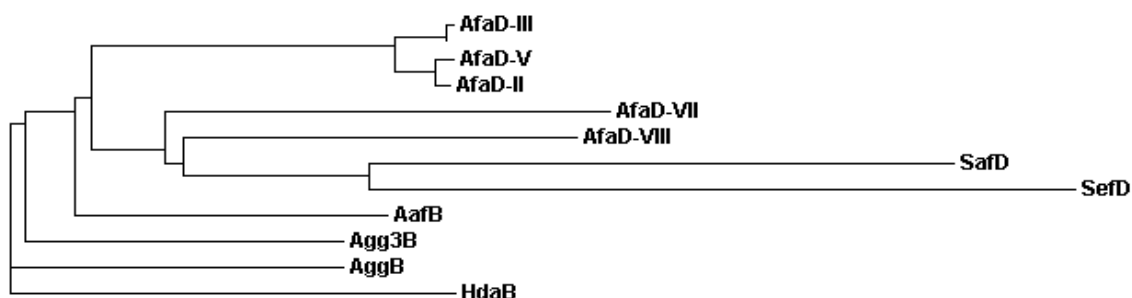


Figure 4.1: Phylogram based on the alignment of AfaD-III homologues. Similarities range from 20% to over 90%.

4.1.1 Designing of Constructs

Both the major subunit (adhesion) and minor subunit (invasion) has been engineered to relocate the extended β -strand from the N-terminus to the C-terminus (DSC strand), to mimic the intermolecular interaction that occurs *in vivo*, to prevent polymerisation and to strengthen the stability of the monomeric form of the protein (Fig 4.2). Furthermore, a DSC strand was also designed to link the two subunits together to form a dimer (see Material and Method for details on the constructs).

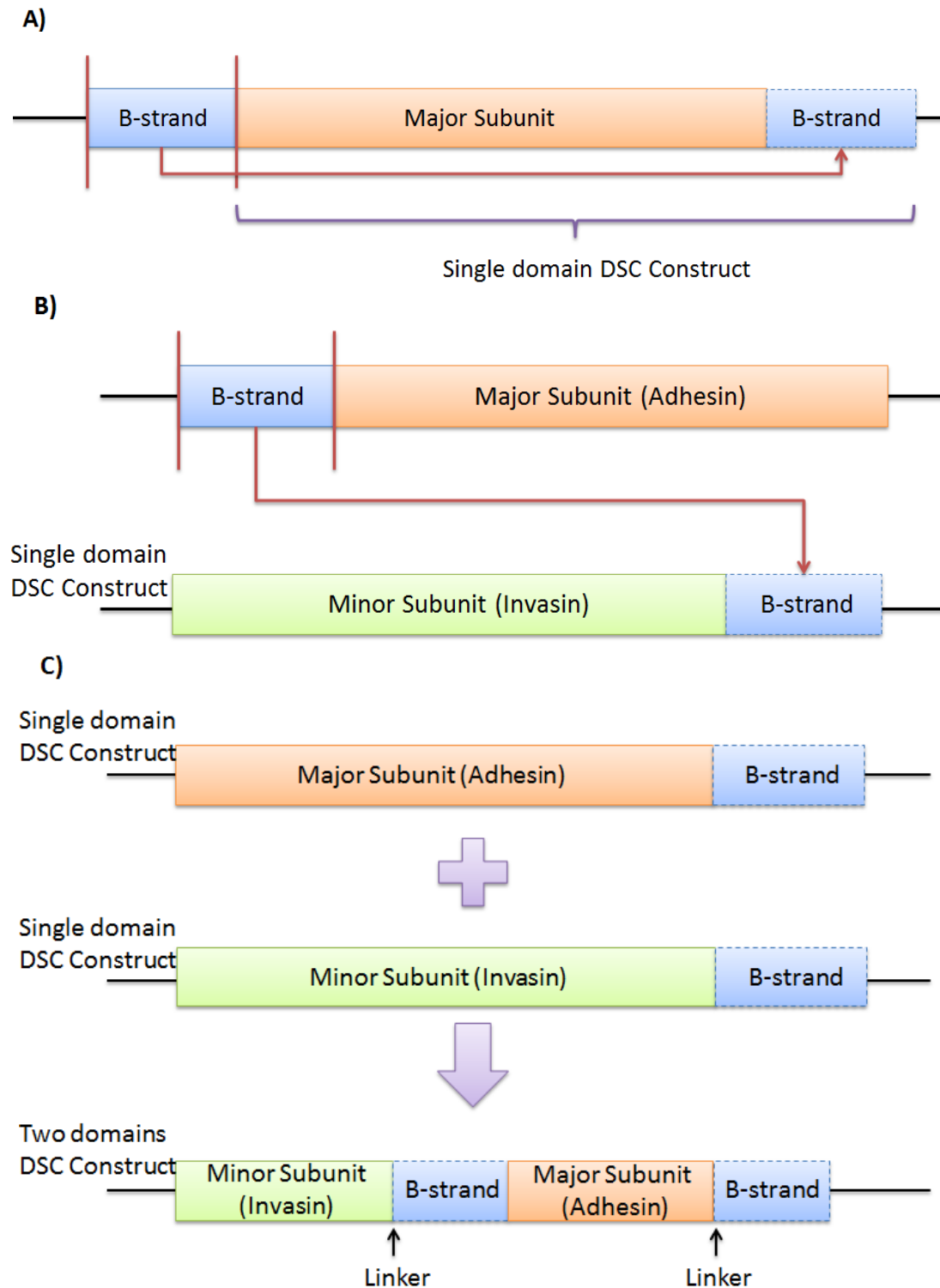


Figure 4.2: A schematic diagram demonstrating the design of constructs. A) The DSC construct of adhesins were constructed by relocating the β -strand from the N-terminus to the C-terminus. B) The invasin DSC constructs were built by adding the β -strand sequence from the adhesin to its C-terminus. C) The two domain DSC constructs were engineered by linking the two single domains DSC constructs together.

4.1.2 Polymerase Chain Reaction

The constructs of the dimer of minor subunit and major subunit were synthesised by Eurofins where each minor subunit and major subunit was subcloned out by polymerase chain reaction (PCR) and was ran on a 1.2% agarose gel (Fig. 4.3) and displayed the correct sizes (~500 bp). The *aafA* construct was provided by Dr. Y. Yang (Imperial College).

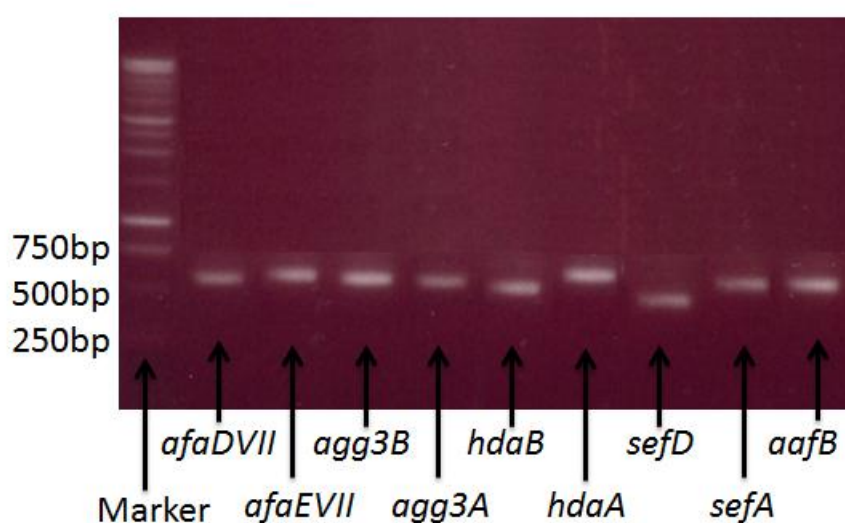


Figure 4.3: An image of 1.2% agarose gel confirming a succesful PCR of the constucts.

4.1.3 Ligation to PQE-30

As with the design all of these constrcuts each contained resitrcition sites, *BamHI* and *HindIII* on 5' and 3' end respectively. Therefore, both the inserts and the plasmid (PQE-30) were digested with both enzymes to obtain compatible ends for ligation. After ligation, each construct was transformed to XL1-Blue competent cells and DNA was harvested from transformation. The identity of each clone was confirmed by DNA sequencing.

4.2 Expression and Purification

After confirming the DNA sequence of each construct, a transformation to M15 *E. coli* cells with plasmid *pRREP4* was performed for protein expression. As described in Materials and Methods (Chapter 3), proteins were purified under denaturing conditions (8 M urea) and refolded by dialysis. Expression of these proteins is indicated in Fig. 4.4. Expression was unsuccessful with AfaE-VII, AfaDE-VII, Agg3A, Agg3B and Agg3BA; therefore no further studies were carried out for these proteins.

After purification with Ni-NTA agarose, some samples still contained impurities (as indicated by SDS PAGE (Fig. 4.4) and further purification with gel filtration was necessary. Both HdaA and HdaB displayed two clear bands of the molecular weight similar to their monomeric and dimer form (Fig. 4.4). Similarly, for HdaBA two distinctive bands and some smearing were observed. The strong and weak bands correspond to the monomeric and dimeric forms of the protein, respectively, and the smear suggested potential polymerisation event but could arise from impurities. In addition, two bands were observed with the SefA sample which indicated the presence of both monomer and dimer.

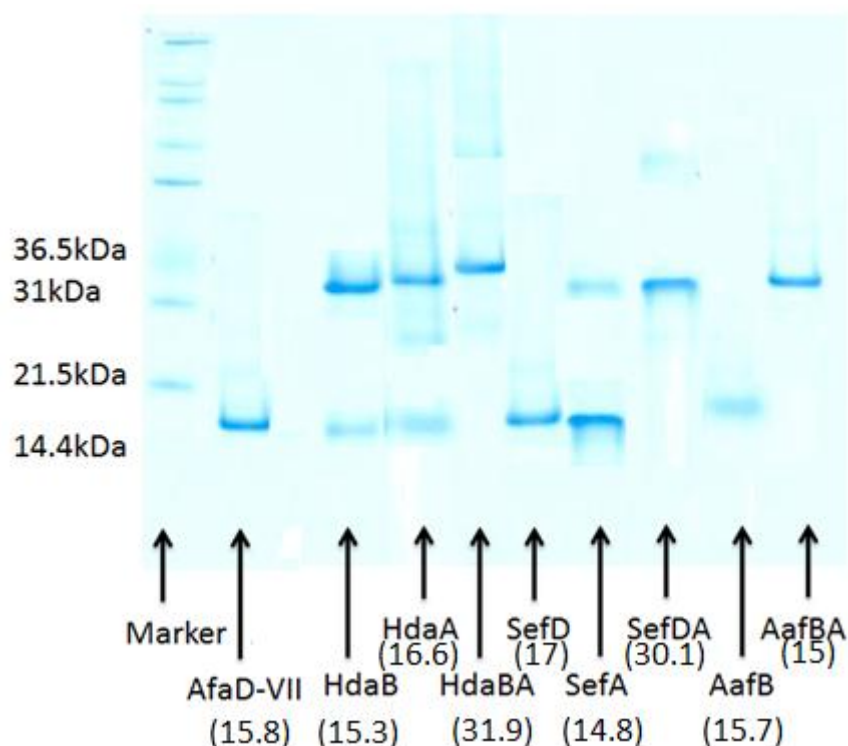


Figure 4.4: A SDS-PAGE showing samples from protein expression. The molecular weight of the marker is given on the left and each lane was labelled with the protein sample.

4.2.1 Size Exclusion Chromatography

All the protein samples were concentrated to 1 ml and passed through a Superdex-75 prep grade column with pH 5 50 mM acetate buffer and 200 mM sodium chloride (Fig. 4.5).

The sample of AafB displayed a chromatogram with one intense peak of expected molecular size of a monomer (Fig. 4.5). The sample from the peak was collected for crystallisation trials and the samples from the small aggregation peaks were discarded.

An intense peak at expected elution volume of the monomer of SefDA exhibited a “shoulder”. The protein seen at the dimer molecular weight (weak) on SDS-PAGE is likely to lie within this shoulder and represent impurities. As it was difficult to separate the dimer, only the monomer was collected for further studies.

A single peak at the expected molecular weight of a monomer was observed for the SefD sample. Aggregates were also present, indicated by a shoulder the peak. Therefore, a small volume at the beginning of the peak was discarded to prevent contamination. Other samples clearly demonstrated multiple peaks in the chromatography and will be described further in the chapter.

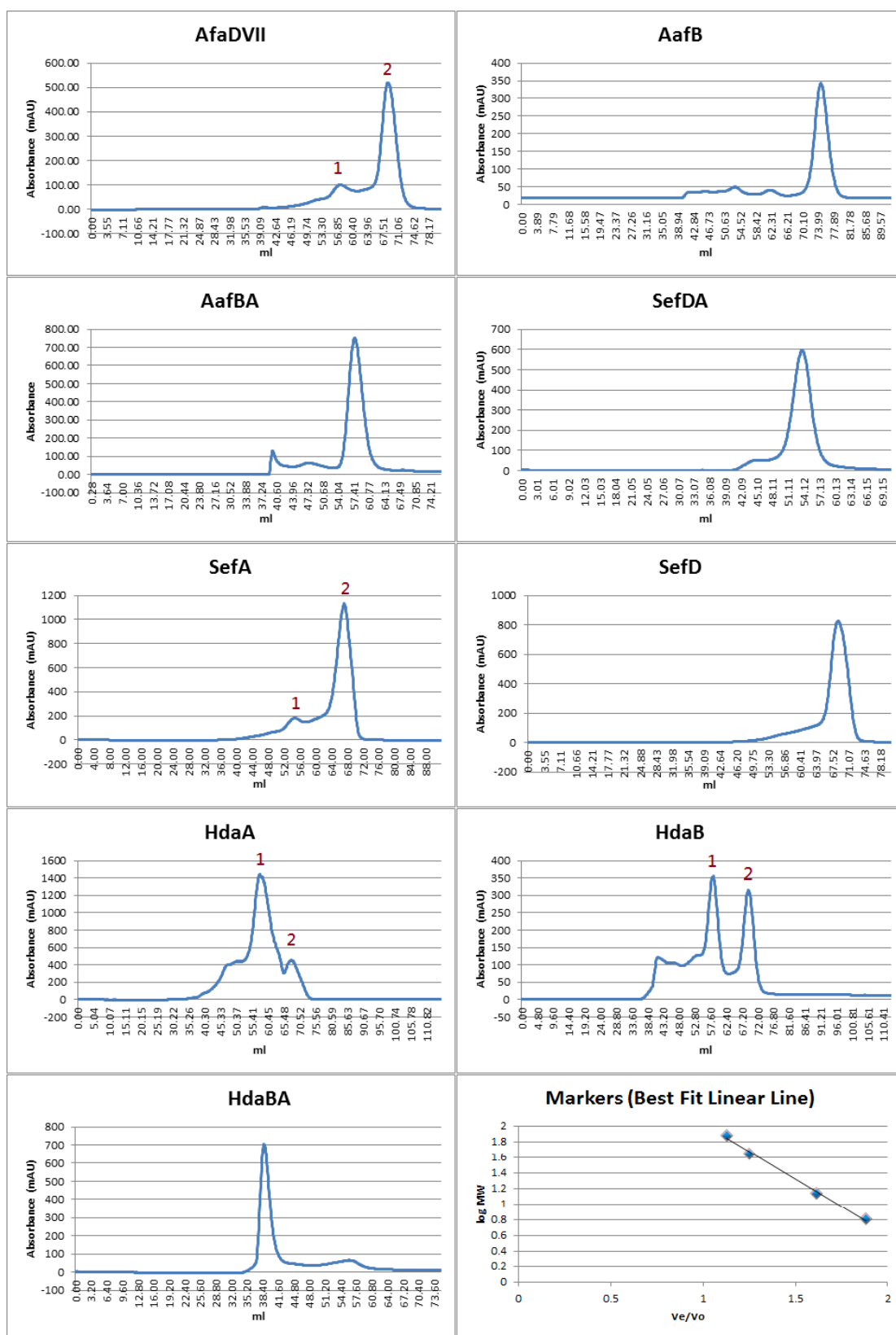


Figure 4.5: The results of Size exclusion chromatography. The sample of intense single peak was collected (AafB, AafBA, SefDA, SefD and HdaBA) and two samples were collected for multiple peaks as numbered 1 and 2 (AfaD-VII, SefA, HdaA and HdaB).

4.2.1.1 AfaD-VII

In the case of AfaD-VII, two peaks were observed which indicated the presence of two species, where the most intense one corresponds to the monomeric form of the protein (15.8 kDa) and the weaker peak displayed the molecular weight expected of a dimer. Samples of both peaks were collected and run on a SDS-PAGE where both lanes showed a single band with identical size consistent with the monomer (Fig.4.6). This suggested a small portion of the AfaD-VII molecules were forming dimers possibly by disulphide bond, but the intermolecular interaction might not be strong hence the dynamics was favoured toward the monomeric form. However, as the SDS-PAGE confirmed two samples were likely to be AfaD-VII both samples were used for crystallization trials.



Figure 4.6: A SDS-PAGE of the samples collected from the two peaks observed in size exclusion chromatography. Both lanes show bands corresponding to the molecular weight of a monomer.

4.2.1.2 AafBA

An intense peak was observed with the expected molecular weight of the monomeric form of AafBA. However, another small peak and a shoulder were also present which

were likely to arise from aggregates of the protein as the position of this peak was beyond the size limit of the column (Superdex-75). As the shouldering was very close to the main peak, samples were carefully collected by discarding some of the volume at the beginning of the peak.

4.2.1.3 SefA

Both SDS-PAGE and size exclusion chromatography suggested the present of a dimer as well as the monomer, so samples from both peaks were collected. These samples were run on a SDS-PAGE and displayed an intense band for monomer and a much weaker band for the dimer. This suggested that the monomeric form was more favourable and the intermolecular interaction might cause by the misfolding of disulphide bond. Nevertheless, both monomer and dimer samples were collected for crystallization trials.

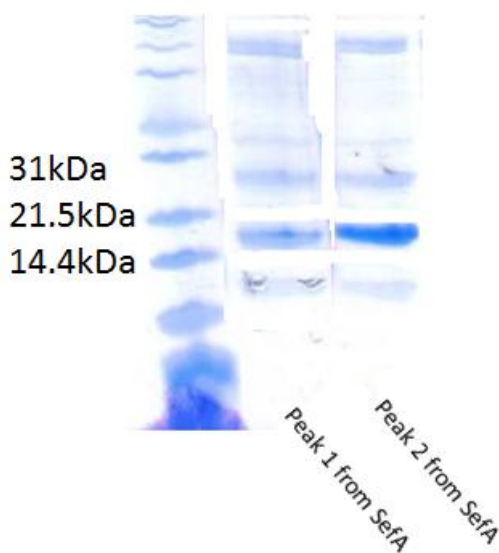


Figure 4.7: SDS-PAGE of the SefA samples from size exclusion chromatography. Both samples shown similar bands which the most intense one corresponding to the size of the monomer. The samples were loaded onto the same SDS-PAGE and was cut and merged for the purpose of this image.

4.2.1.4 HdaA

In size exclusion chromatography of HdaA two peaks were seen, where the intense one is consistent with the molecular weight of a dimer and the later (weaker) peak was at elution volume expected of the monomer. Samples from both peaks were collected. SDS-PAGE revealed a band at the expected molecular weight of the dimer which was stronger than the monomer band (Fig. 4.8). These suggested that the dimer form is more favourable as expected to be the result of the designed additional disulphide bond. Although the original aim was to produce a stable monomer, this dimer seems to be strong and stable. Both samples were collected for further study to reveal the true cause of the intermolecular interaction.

4.2.1.5 HdaB

In size exclusion chromatography three peaks were observed, where the first peak was at the position beyond the size limit of the column (Superdex-75) so most likely arise from impurities or aggregates and was discarded. As the other two peaks were both sharp and corresponding to the molecular weight of a dimer and a monomer respectively, both samples were collected and ran on SDS-PAGE (Fig. 4.8). Both samples displayed two bands of the size of dimer and monomer, similar to HdaA. As this intermolecular interaction was also present in HdaB it was suspected to be an artefact due to our construct design. Both the monomer and dimer samples from the size exclusion chromatography were collected for further structural determination.

4.2.1.6 HdaBA

A relatively clean peak was observed for HdaBA, but because the position of the peak was beyond the size limit of the column (Superdex-75) the sample was run on SDS-

PAGE (Fig. 4.8). There was a strong band of the expected molecular weight of a monomer and a weak band of consistent with dimer similar to the one displayed after the protein purification. This suggested a possible of a multimer of HdaBA but the intermolecular interaction might not be strong enough to keep it in place as the result of the dimer band was weak and smeared. As this multimer of HdaBA was beyond the size capacity of the chromatography column, it was difficult to determine its molecular weight but SDS-PAGE indicated HdaBA was present so it was collected for crystallization trials.

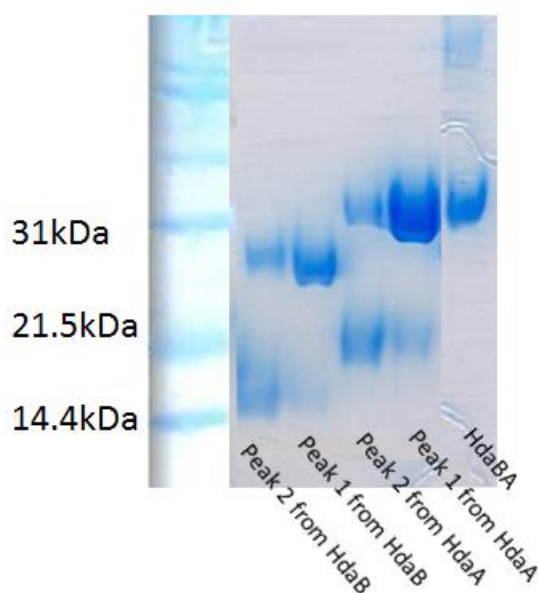


Figure 4.8: SDS-PAGE of HdaA, HdaB and HdaBA samples collected after from size exclusion chromatography. The SDS-PAGE shows two bands for each sample which correspond to the size of dimer and monomer. The samples were loaded onto the same SDS-PAGE which was cut and merged for the purpose of this image.

4.3 One-Dimensional NMR

As all our proteins were purified under denaturing conditions and refolded by dialysis, a portion of each sample was characterised by one-dimensional NMR spectroscopy to assess if the protein was folded. All protein samples were run under the same condition at 298 K and distinct peaks were observed on both the methyl and amide

regions of the spectrum (Fig. 4.9). These chemical shifts were caused by the unique chemical environment of each folded protein.

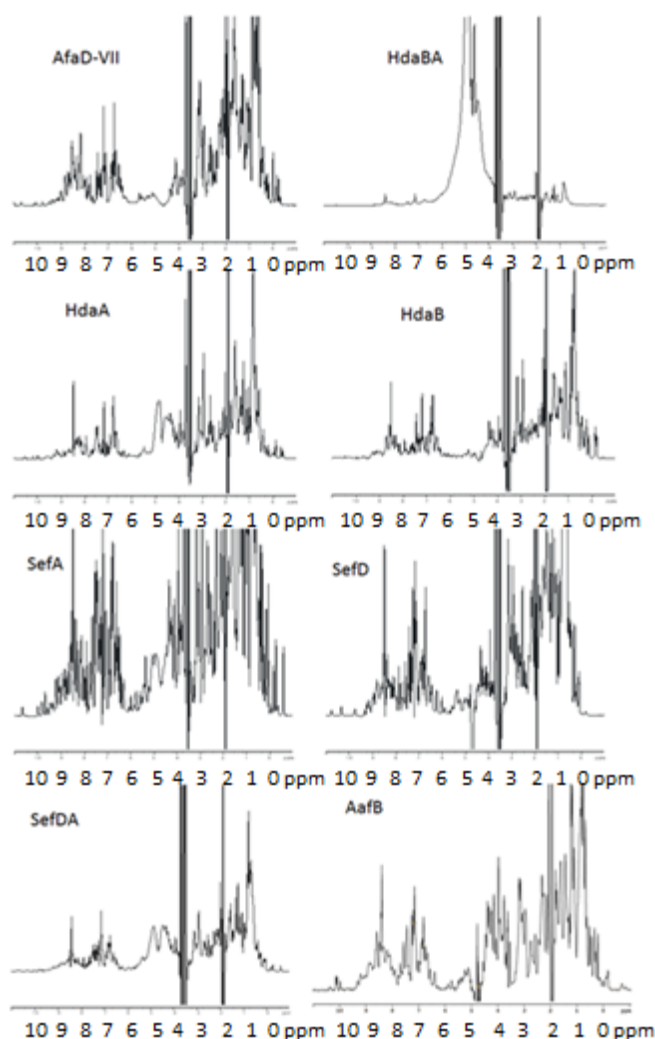


Figure 4.9: One-dimensional spectra of purified protein samples. Most spectra show distinctive peaks in both the methyl and amide region which confirmed the presence of folded protein in all samples. However, the spectra of HdaBA suggested an unfolded protein.

4.4 Crystallisation

Samples from size exclusion chromatography were collected and concentrated to 1 ml for initial crystallisation condition screening.

With AfaD-VII (5.74 mg/ml) the commercial crystallisation condition screens were mostly clear with some light precipitates; therefore further concentration was needed

for crystallisation. Both the monomer (10.58 mg/ml) and dimer (17.12 mg/ml) samples of HdaA seemed to be too low in concentration as the results most of the commercial crystallisation conditions resulted in clear drops; an increase in concentration was also needed. The samples of the SefA dimer (4.22 mg/ml) and monomer (47.77 mg/ml) were tested with commercial crystallisation conditions but only a handful of conditions yielded light precipitates and others showed only clear drops. Furthermore, no crystal formation was observed with the samples of AafBA (8.34 mg/ml), SefDA (7.64 mg/ml) and HdaBA (5.97 mg/ml). However, AafB, HdaB and SefD all showed promising crystals in some conditions and were investigated further.

4.4.1 AafB

After one week many crystals were obtained in various conditions. Due to the high salt concentration in the protein buffer, Izit crystal dye was used to eliminate false hits; the dye can fill up the solvent channels in protein crystals but not salt crystals. Three conditions displayed protein crystals and were used for further optimization (Fig. 4.10). Optimisations were performed by the hanging drop method of vapour diffusion with the conditions given in Table 4.1. The buffer conditions were pH 5, 50 mM acetate with 200 mM sodium chloride with a protein concentration of 7.6 mg/ml. In the optimization step both the pH and the precipitant concentration were altered using a 2-dimensional screening process (Table 4.1). Crystals were observed in all conditions, although most of the crystals were small, the larger crystals were looped-out and a range of cryoprotectants were tested by exposing the crystals to X-rays and assessing the resolution limit and presence of ice rings. A ratio of 1:1, well solution to a mixture of PEG3350 and PEG400 gave the best results.

Protein	Buffer pH	Salt	Concentration of the Precipitant
AafB	pH 6.4-pH 7.8 of 0.2 M DL-Malic Acid	—	18%-23% Polyethylene glycol (PEG) 3350
AafB	pH 3.4-pH 4.8 of 0.1 M Sodium Citrate	1 M Lithium Chloride	18%-23% Polyethylene glycol (PEG) 6000
AafB	pH 4.9-pH 6.3 of 0.1 M BIS-TRIS	0.2 M Lithium sulphate	23%-28% Polyethylene glycol (PEG) 3350

Table 4.1: Range of optimisation conditions for crystallisation of AafB

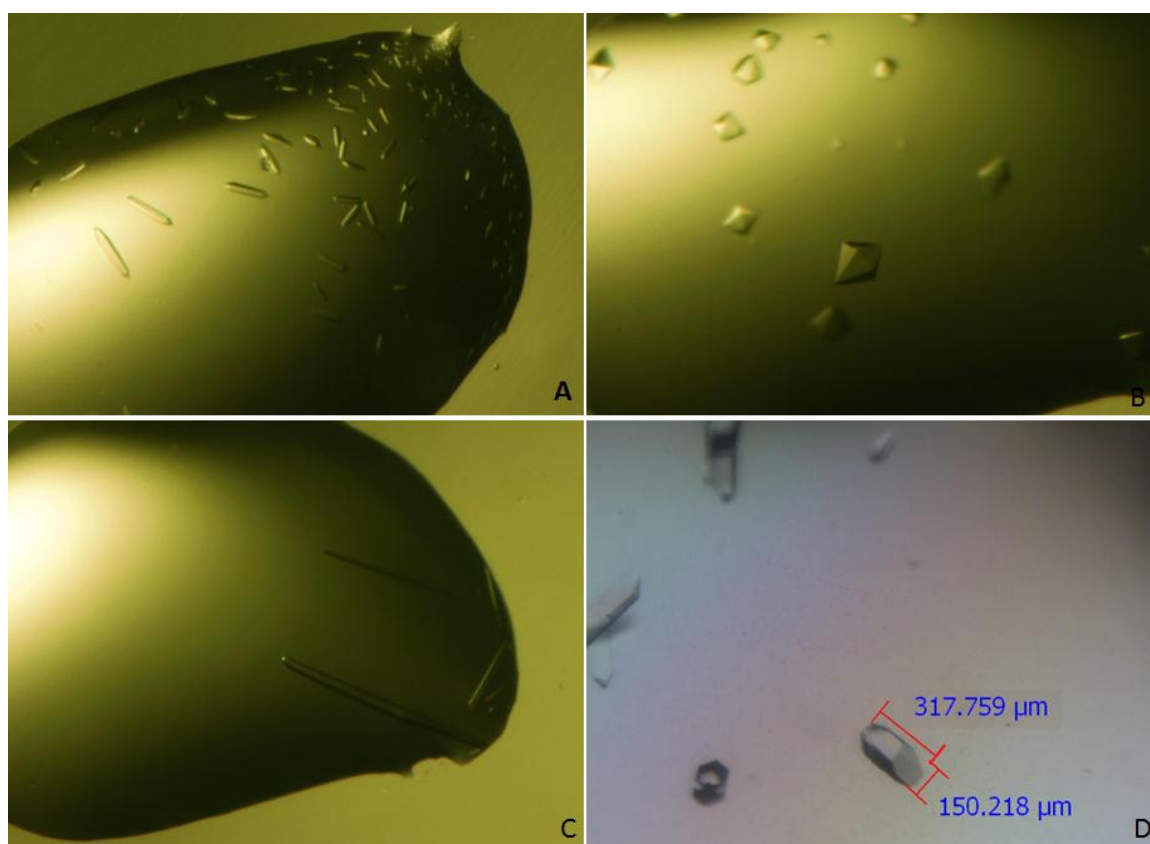


Figure 4.10: Images of crystals observed from crystallization screen. A) Rod shaped crystals of AafB in lithium chloride and 20% PEG 6000, pH 4.0 sodium citrate. B) Diamond shaped crystals of AafB in lithium sulphate and 25% PEG 3350, pH 5.5 Bis-Tris. C) Needle shaped crystals of AafB in DL-malic acid and 20% PEG 3350, pH 7.0. D) A typical crystals for looping out and exposed to X-rays.

4.4.2 HdaB

Two conditions were found to yield crystals for HdaB (6.24 mg/ml) and optimisation was performed around these conditions (Table 4.2). However, only crystals from one set of conditions (Ammonium Citrate with PEG3350) were looped-out and mixed with cryoprotectant and 500 mM sodium iodine, because the other optimization conditions yielded mostly precipitate and some tiny crystals (Fig. 4.11). Furthermore, in order to provide more options for obtaining phase information, some crystals were also soaked in freshly made 8 M sodium iodide with cryoprotectant.

Protein	Buffer pH	Salt	Concentration of the Precipitant
HdaB (pH 5 50 mM Acetate Buffer with 200 mM Sodium Chloride)	pH 4.0-pH 5.4 of 0.1 M Sodium Acetate	0.2 M Ammonium Sulphate	25%-30% Polyethylene glycol (PEG) 550 MME
HdaB (pH 5 50 mM Acetate Buffer with 200 mM Sodium Chloride)	pH 4.5-pH 5.9 of 0.2 M Ammonium Citrate	-	18%-23% Polyethylene glycol (PEG) 3350

Table 4.2: A range of optimisation conditions for crystallisation of HdaB.

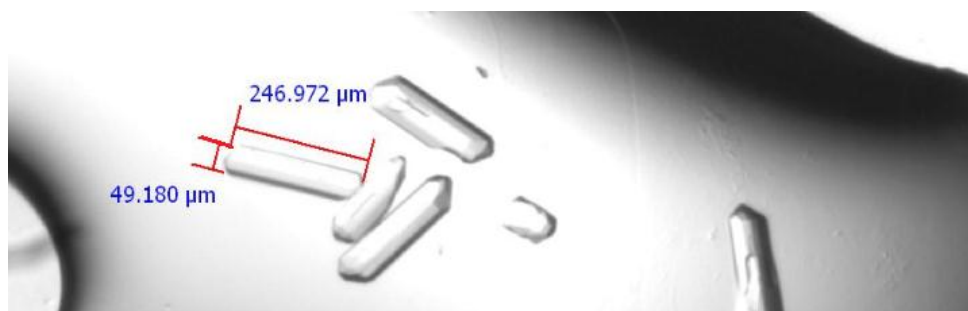


Figure 4.11: A typical drop with protein crystals from the crystallisation condition of ammonium citrate with PEG 3350.

4.4.3 SefD

The protein buffer remained the same in SefD (pH 5 50 mM acetate Buffer with 200 mM sodium chloride) and crystals were found in four conditions in the initial screen

(Table 4.3). In addition, due to the high salt condition, Izit dye was used and confirmed the presence of protein crystals (Fig. 4.12). Amongst these conditions only one (sodium acetate with ammonium acetate) produced crystals of workable size but they could not be reproduced. A micro-seeding method was used but the crystals remained small in size and clustered together and were difficult to loop-out for further experiments. A commercial additive screen was also used, but most conditions resulted in precipitation or clear drops. A few conditions resulted in small crystals which were unsuitable for further study. Nevertheless, big crystals were produced by changing the crystallisation buffer from H₂O to D₂O as discussed in Chapter 8. Some crystals were looped-out and passed through cryoprotectant but some crystals were also soaked in freshly made sodium iodide and 5-amino-2, 4, 6-triiodoisophthalic acid (I3C) with cryoprotectant. Various concentrations of sodium iodide and I3C were tested and a concentration of 2 M was the best condition as it allowed soaking without any noticeable damage to the crystals.

Protein (7.18 mg/ml)	Buffer pH	Salt	Concentration of the Precipitant
SefD	pH 4.0-pH 5.4 of 0.1 M Sodium Acetate	—	1.3%-1.8% Lithium Sulphate
SefD	pH 4.6-pH 6.0 of 0.1 M Sodium Citrate	0.1 M Magnesium Chloride	12%-17% Polyethylene glycol (PEG) 4000
SefD	pH 6.2-pH 7.6 of 0.1 M Succinic Acid	-	18%-23% Polyethylene glycol (PEG) 3350
SefD	pH 4.0-pH 5.4 of 0.1 M Sodium Acetate	-	3.7 M-4.2 M Ammonium Acetate

Table 4.3: A range of optimisation conditions for crystallisation of SefD.

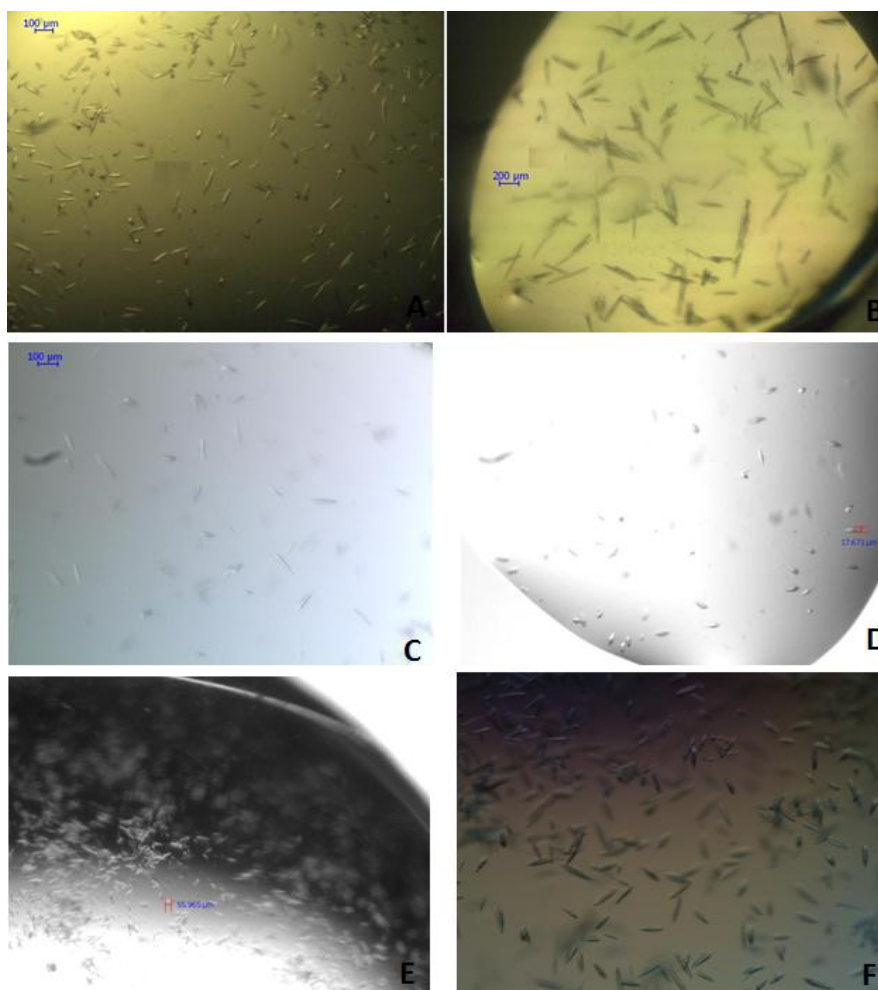


Figure 4.12: Images of crystals observed from different crystallization conditions and methods. A) Tiny crystals forming from Sodium Acetate and Lithium Sulphate. B) Clustered of thin needle like crystals observed from the condition of Sodium Citrate, Magnesium Chloride and PEG4000. C) Thin needle like crystals found in the condition of Succinic Acid and PEG3350. D) An example of tiny crystals formation from additive screens. E) A typical image of seeding experiments showing the present of tiny crystals. F) Images taken after Izit dye experiments which confirmed of genuine protein crystals.

Chapter 5: Structural Determination and Analysis of AafB

5.1 Data Collection

Crystals of AafB were quickly passed through cryoprotectant (the composition of the cryoprotectant is given in section 3.4.3) and then flash frozen in liquid nitrogen. These were taken to beamline I24 (Diamond Light Source, UK) and exposed to synchrotron radiation with wavelength 0.9778 Å. The smaller crystals showed diffraction to around 9 Å whilst some of the larger crystals diffract to higher resolution of 3.2-2.95 Å. The best dataset was obtained from a large (50x50x150um³) well cryoprotected crystal which diffracted beyond 3.0 Å. After indexing 2 images separated by 90°, the point group was suggested to be P6₂2 which only requires 30° of data. However, in order to provide more data for any unforeseen problems later in the processing, 180° of data were collected. The dataset was collected at 100 °K with an oscillation angle of 0.2° per frame to 2.5 Å on a Pilatus 6M detector, with a detector distance of 527.77 mm.

5.2 Data Processing

The data were first indexed in the lower symmetry point group P3 to 3.0 Å using MOSFLM (1994) and scaled with SCALA (1994). The results from SCALA were fed into Pointless (1994) which confirmed the point group of P6₂2 and suggested a space group of either P6₁22 or P6₅22 based on the strong systematic absences. Therefore, the data was reprocessed, scaled and truncated in space group of P6₁22. The statistics for processing and scaling of the AafB data are shown in Table 5.1. Analysis of the crystal composition gave a Matthews coefficient (*V_m*) of 3.41 and 2.27 indicated that there are two or three protein molecules in the asymmetric unit with a solvent concentration of 63.93% and 45.90% solvent concentration, respectively (Table 5.2).

Protein		AafB (with AafA DSC strand)
Space group		P6 ₁ 22
Cell dimensions (Å)	a	81.7553
	b	81.7553
	c	222.7027
Cell angles (°)	α	90.00
	β	90.00
	γ	120.00
Number of reflections (total)		58800
Number of reflections (unique)		9477
I/sigI (mean)		4.8 (2.0)
Multiplicity		2.9 (2.4)
R _{sym} (%)		11.1(37.4)
Completeness (%)		97.0 (89.4)
Resolution range (Å)		50.802 - 3.0 (3.16 - 3.00)
Average Wilson B Value (Å ²)		74.155

Table 5.1: Summary of data collection and scaling statistics for AafB. Statistics for the outliers are in bracket.

Number of molecules per Asymmetric unit	Matthews Coefficient (Vm)	Percentage of Solvent Content
2	3.41	63.93
3	2.27	45.90

Table 5.2: Summary of Matthews Coefficient of AafB. The number of molecules per asymmetric unit is given with the corresponding Matthews coefficient score and solvent content.

5.3 Structural Determination

Phase information was calculated in PHASER (1994) by molecular replacement using residues 1 to 118 from chain A of the published DraD crystal structure (PDB: 2AXW (Jedrzejczak *et al.* 2006)) as the search model. All the anisotropic B factor parameters and hydrogen atoms were removed from DraD using PDBSET (1994) . Molecular replacement was performed between resolution ranges 50.802-3.0 Å and only two molecules were found in the asymmetric unit with the translational Z-scores of 10.1 and 6.8 and the log likelihood gain of 97 and 332, respectively (Table 5.3). Rigid body refinement was performed in PHASER resulting in an R-factor of 0.52. Noticeably, the translational Z-score for the second molecule was lower which was unusual but the electron density showed good connectivity for this region.

	Translational Z- Score	Packing Check (PAK)	Log Likelihood Gain
First Molecule	10.1	0	97
Second Molecule	6.8	0	332

Table 5.3: Z-score of the two molecules found from PHASER. Only two molecules were found in the asymmetric unit with good Z-score and log likelihood gain of 97 and 332 respectively. Packing check confirmed no clashes between the asymmetric units within the unit cell.

The asymmetric unit of AafB consists of two molecules (Fig. 5.1) and the crystal packing forms a well-connected lattice (Fig 5.2) with good connected density for both protein molecules. In addition, the maps demonstrated rounded density for proline residues and extended density for large side chains such as histidine, arginine, glutamic acid, glutamine and asparagine (Fig. 5.3).

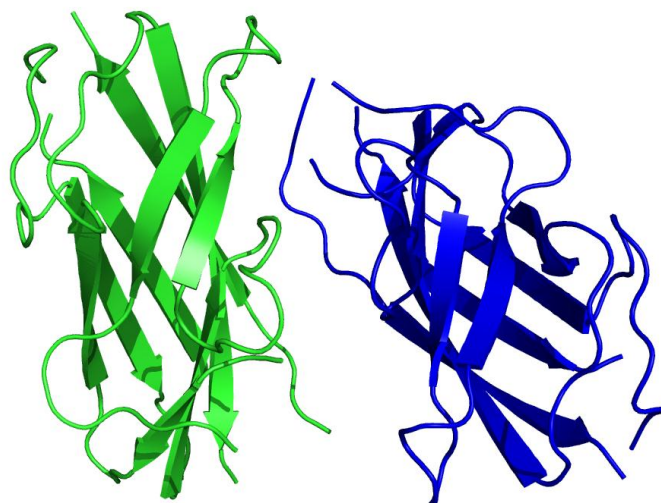


Figure 5.1: The asymmetric unit of the AafB crystal. The AafB asymmetric unit is made up of two molecules which are coloured green and blue.

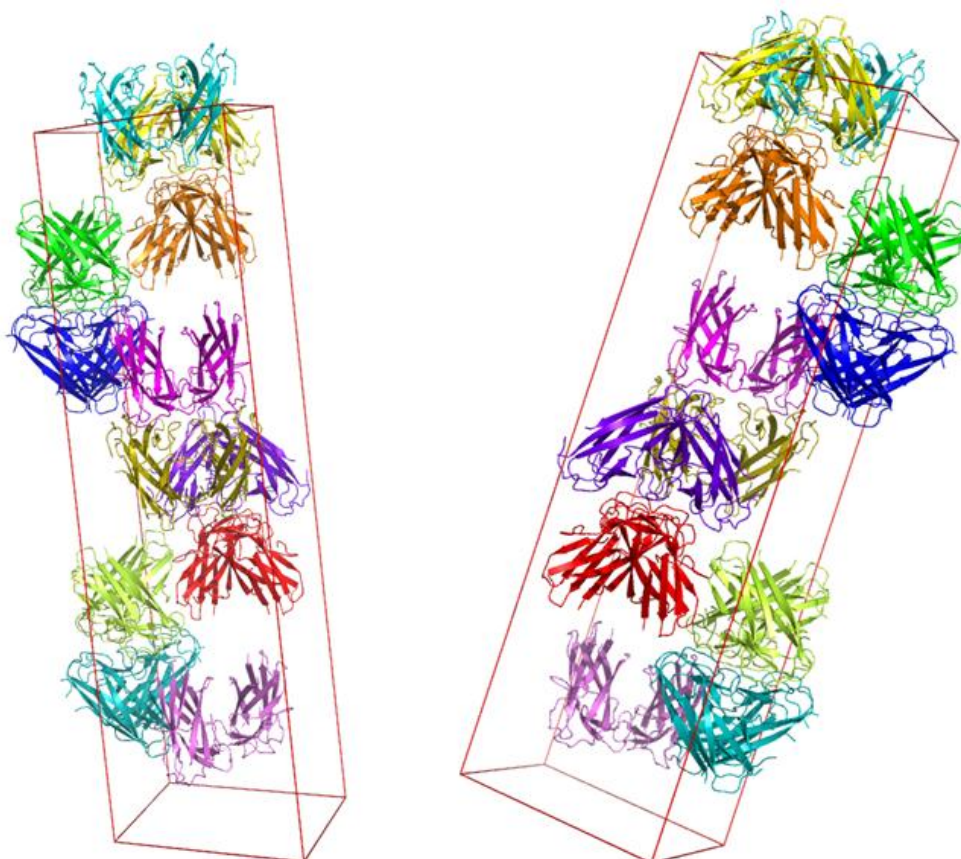


Figure 5.2: Two different views of the crystal packing arrangement within the AafB crystal. There are 12 asymmetric units of AafB in the $P6_122$ unit cell and each are coloured in a different colour.

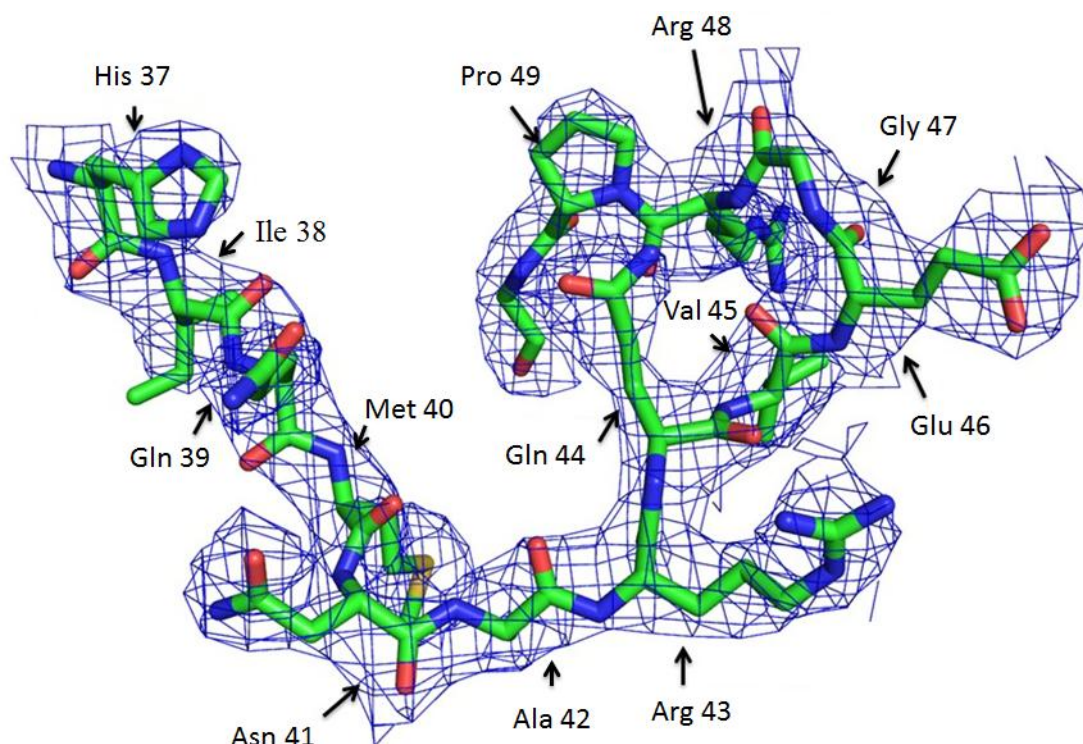


Figure 5.3: An example of the electron density map quality of AafB. The region shown is from residue 37 to 49 of chain A. The backbone is nicely connected with good electron density. The large side chains such as arginine, histidine, glutamine, asparagine and glutamic acid were easily recognizable as shown in the $2F_o - F_c$ electron density map contoured at 1.0 rms electron density. Image was generated by PyMOL.

5.4 Structural Refinement and Modelling

Positional and individual isotropic B-factor refinement were performed with the CCP4 programme REFMAC (1994). Coot was then used to mutate residues to the correct side chains, and make adjustments to the model. This was followed by cycles of refinement in REFMAC using automated 2-fold non-crystallographic symmetry (NCS) restraints, jelly-body restraints and a map-sharpening algorithm, followed by model building in COOT. With automated NCS, REFMAC compares the root-mean-square distance (rmsd) of the molecules in the asymmetric units and applies weights to the atoms. Those with very similar rmsd values are tightly NCS restrained whilst those with very high rmsd values are not restrained. Jelly-body restraints use similar known structures from the REFMAC database to define rigid body restraints to reduce

the gap between R_{work} and R_{free} in low resolution data. Finally the map sharpening function increases the signal to noise ratio and hence can greatly improve the electron density map at low resolution.

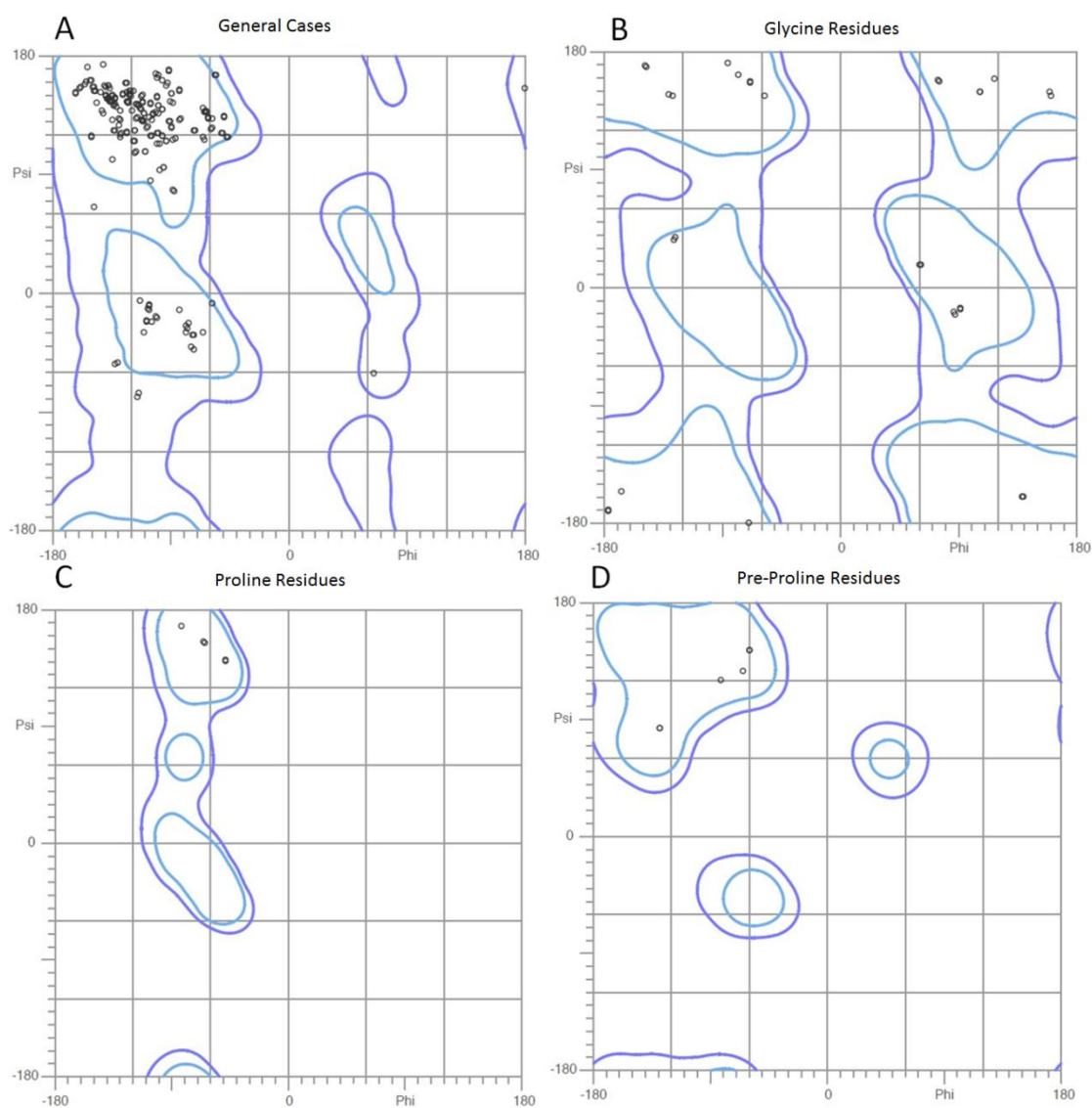
With the resolution of this dataset (3.0 Å) only 5% of the data was removed for cross-validation during refinement. The model was refined using the 95% working data set with the aim of minimizing the value of R_{work} which was obtained after each round of refinement. To evaluate the accuracy of the model and protect against model bias, the smaller dataset was used to calculate the R_{free} and was expected to correlate with R_{work} . As in this case, the model had similar R_{work} and R_{free} which suggested no bias in the model and with no possible adjustment to the protein model, based on inspection of the electron density maps and no further reduction in the R-factors, a search of ions and water molecules was performed. This resulted in two water molecules, one sulphate ion and one acetate ion. The statistics of the final refinement as given in Table 5.4.

Refinement Statistics		Values
Refinement Resolution (Å)		50.802 – 3.00
Reflection used in Refinement		8638
Reflections used in R_{free} Calculation		437
Total Protein Residues in Asymmetric unit Model		267
Number of Protein Molecules in the Asymmetric unit		2
Number of Ions		2 (one sulphate ion and one acetate ion)
Number of Water Molecules		2
Number of Non-hydrogen Atoms		1933
RMS Deviation from Ideal	Bond Length (Å)	0.009
	Bond Angles (°)	1.181
	Dihedral Angles (°)	6.091
R_{work}		0.22149
R_{free}		0.24840

Table 5.4: Refinement statistics for the final refined AafB model.

5.5 Model Analysis and Validation

With no further refinement or model building leading to improvements to the model, the structure was validated using COOT, MOLPROBITY (Chen *et al.* 2010) and the PDB server. All atoms in the backbone and side chains were surrounded by good electron density and the orientation of each side chain was adjusted by rotamer analysis. With the low resolution of the data, both the bond lengths and bond angles were tightly restrained during refinement to those of theoretical values. Furthermore, no additional free electron density blobs were discovered at this stage which suggested no additional ligand or water molecules were present. The Ramachandran analysis reported by MOLPROBITY gave 96.8% (239/247) of the residues resident in the favoured regions and just 3.2% (8/247) of the residues falling within the allowed regions (Fig. 5.4).



Residues in most favourable regions	239	96.8%
Residues in allowed regions	8	3.2%
Residues in disallowed regions	0	0%
Number of non-glycine, non-proline and pre-proline residues	221	100%
Number of glycine residues	18	
Number of proline residues	4	
Number of pre-proline residues	4	
Total number of residues	247	

Figure 5.4: Ramachandran plots for the structure of AafB. The areas inside the light blue lines represent the favoured regions and the areas within the purple blue regions represent the allowed regions. Image generated from MOLPROBITY (Chen *et al.* 2010).

The final model contains two protein chains, one sulphate ion, one acetate ion and two water molecules. The N-terminal His tags of AafB were not visible in the electron density maps and chain A extends from residue 0 to residue 143 (although residue 10, 77 to 81 and 121 to 129 were not modelled), whilst chain B also extends from residue 0 to residue 143 (with residues 9, 10, 78 to 80 and 122 to 126 absent). These missing residues were part of various flexible loops and were adjacent to residues with high B factors (Fig. 5.5). The overall Wilson B-factor for AafB is 70.04 Å² and correlates well with the final factors for the final structure.

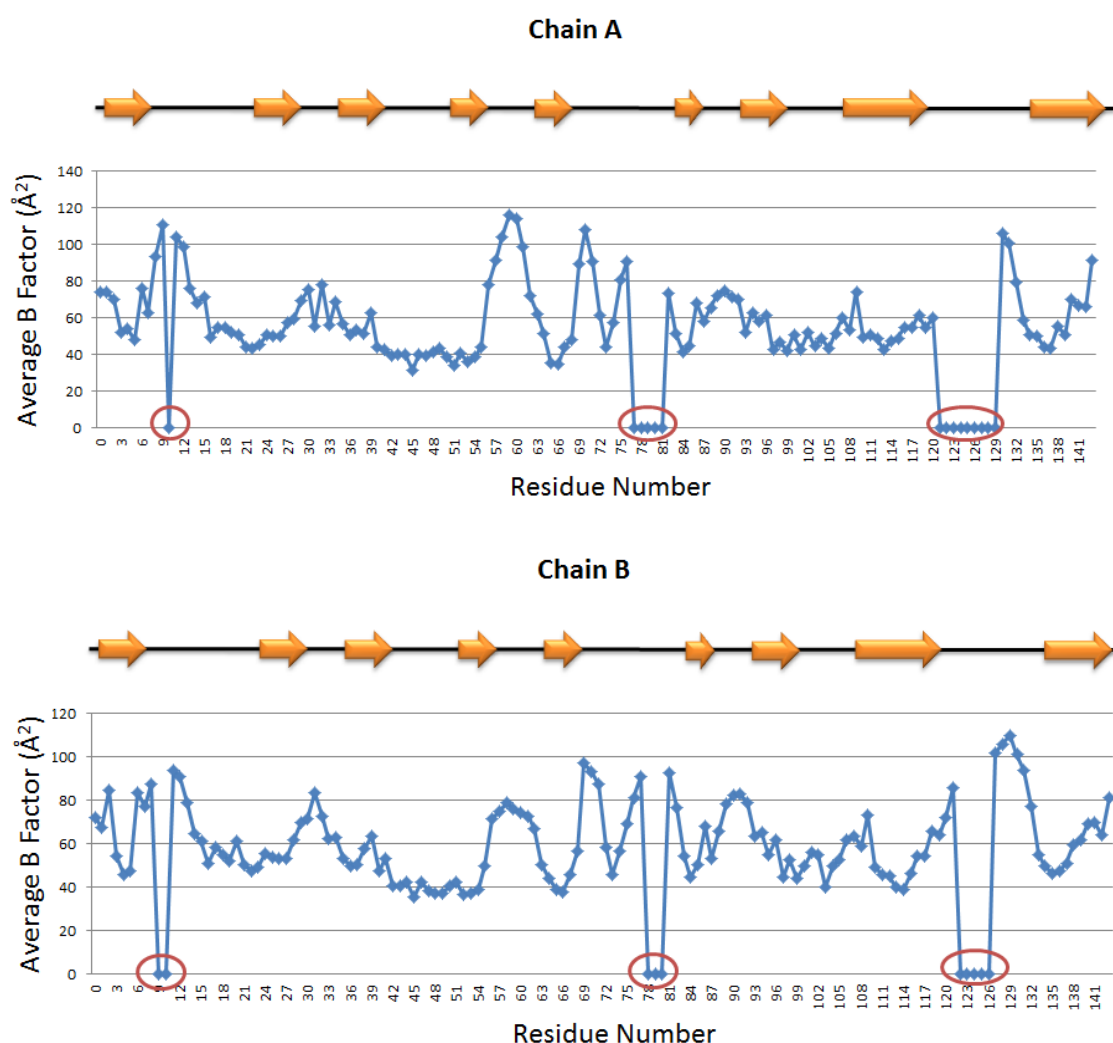


Figure 5.5: Average B factor of AafB residues. The average B factors were plotted for both chains with the value of average B factor on the Y-axis and the residue number on the X-axis. The red circled areas indicate the missing residues and the black lines and the orange arrows above represent the secondary structure.

5.6 Structural Overview of AafB

5.6.1 Crystallography Structure of AafB

The secondary structure of AafB displays two layers of β -sandwich which consist of 9 β -strands including the engineered donor strand on the C-terminus (Fig. 5.6). This immunoglobulin-like fold was expected as the homologous proteins show a similar fold.

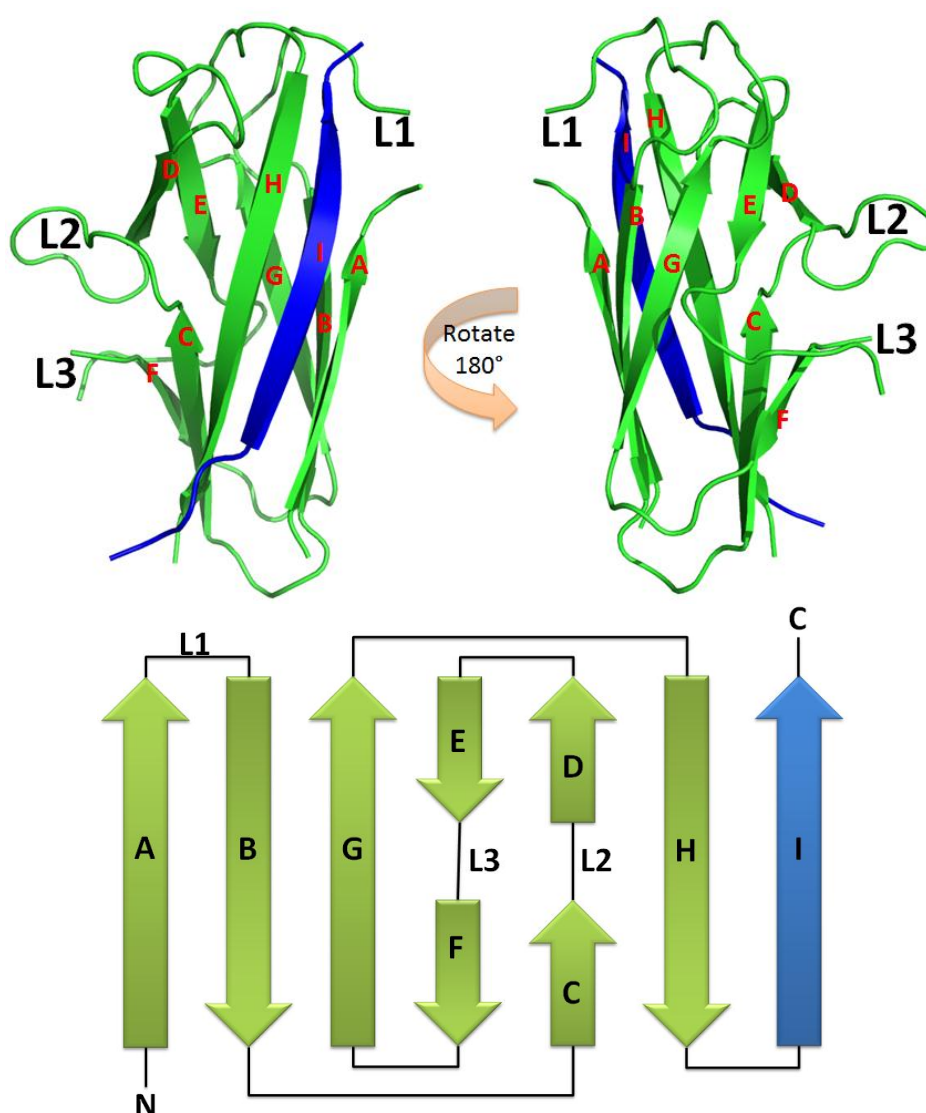


Figure 5.6: Cartoon representation and topology of the Structure of AafB. The top images are the structure of AafB from chain A in the asymmetric unit. The topology representation of the structure is on the bottom with letter labelling corresponded to the order of the strands and the three long loops are labelled L1, L2 and L3. Coloured in green are the native residues and the engineered donor strand is coloured in blue

5.6.2 Comparisons of Molecules within the Asymmetric unit

There are two molecules in the asymmetric unit and they were used to further restrain the rigid body refinement of the structure (to improve the quality of electron density map during processing). Despite the backbone of these molecules being identical (rmsd of 0.137 Å), there were small differences in the loop regions (Fig. 5.7). The three areas with the most disagreement between the two molecules were located where there are missing residues, with adjacent residues displaying high B factors. This suggests high flexibility in these regions due to lack of contact to adjacent molecules (Fig 5.8). In addition, another fully built loop region displays disagreement between the molecules as the result free flexible loop (Fig. 5.8).

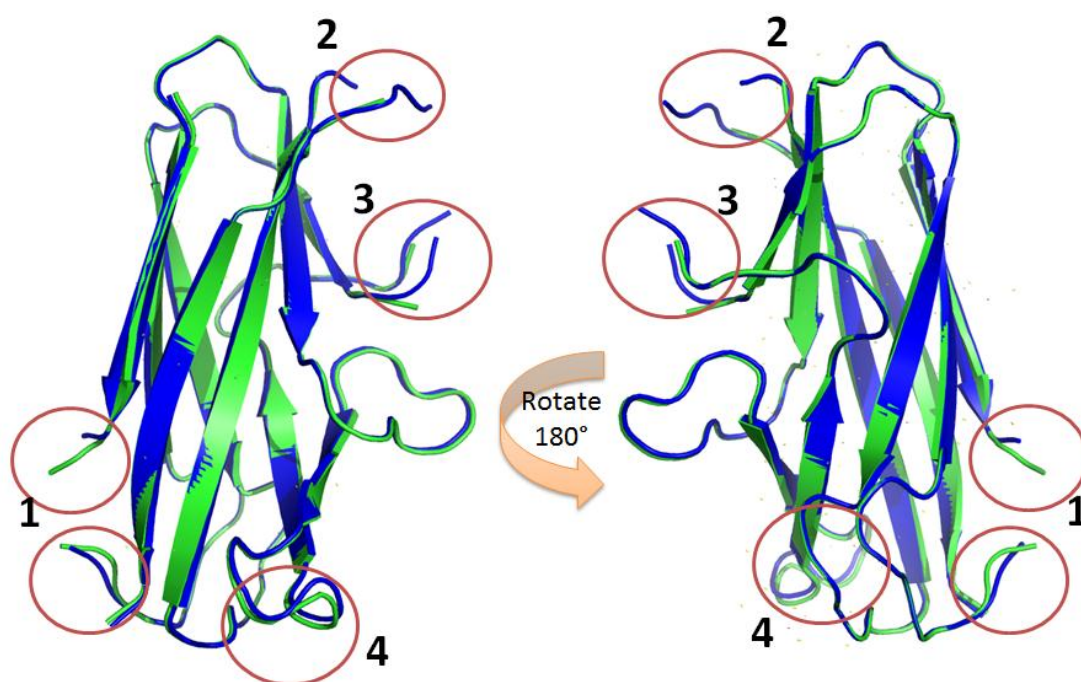


Figure 5.7: Superimposition of the two molecules in the asymmetric unit. One molecule is in green and the other is in blue. The overall structure was very similar with all the β strands nicely overlaid on each other, but differences can be spotted and are labelled number 1 to 4.

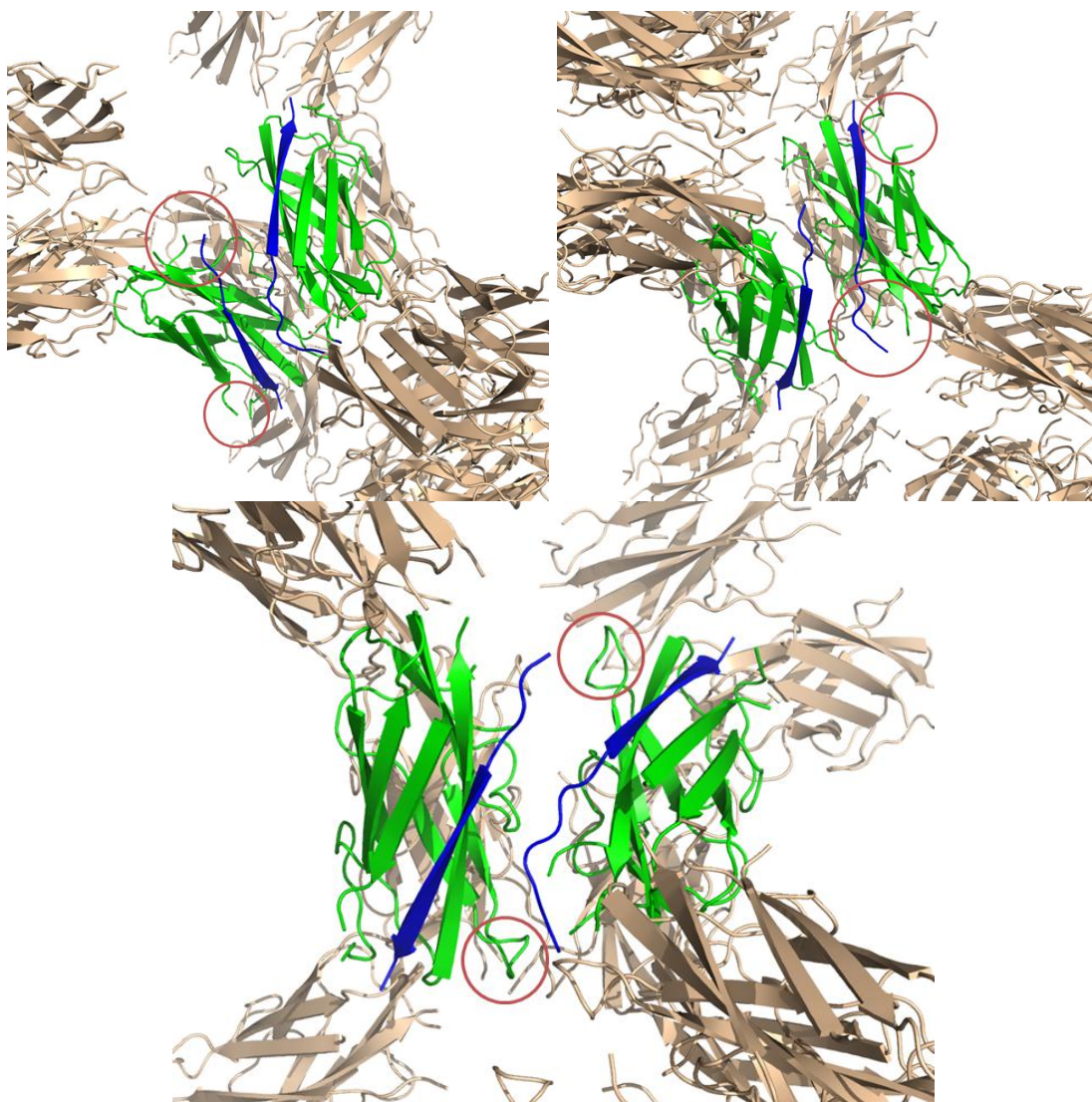


Figure 5.8: Adjacent contacts around the asymmetric unit. The centre asymmetric unit is coloured in green with the donor strand in blue and the adjacent asymmetric units are all in brown. The red circle indicates the loop regions of disagreement where there are no contacts to any adjacent molecules.

Despite the high amount of solvent in the crystal (63.93%), with large solvent channels running through the crystals, each asymmetric unit showed good contacts to adjacent asymmetric units (Fig 5.9).

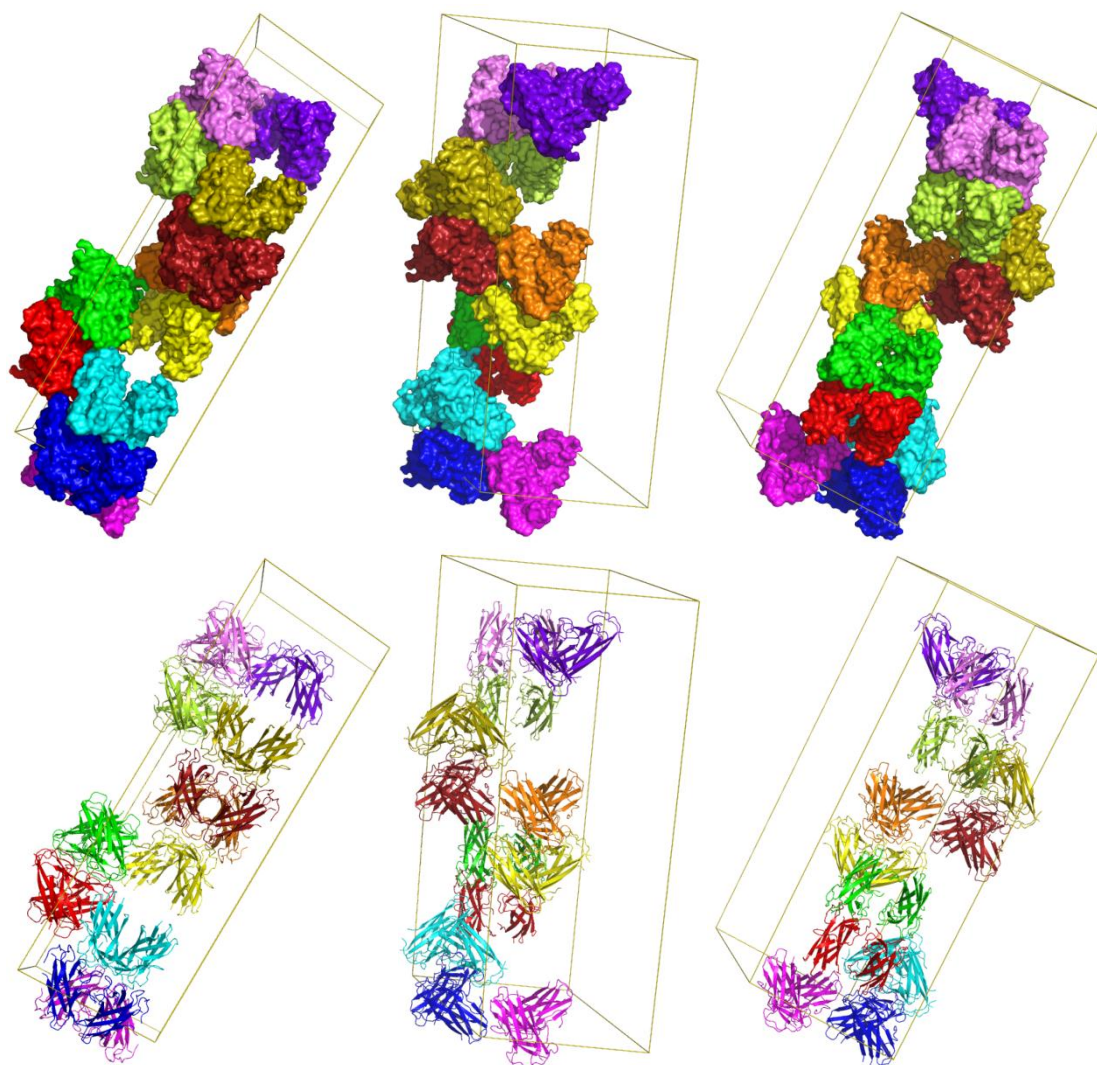
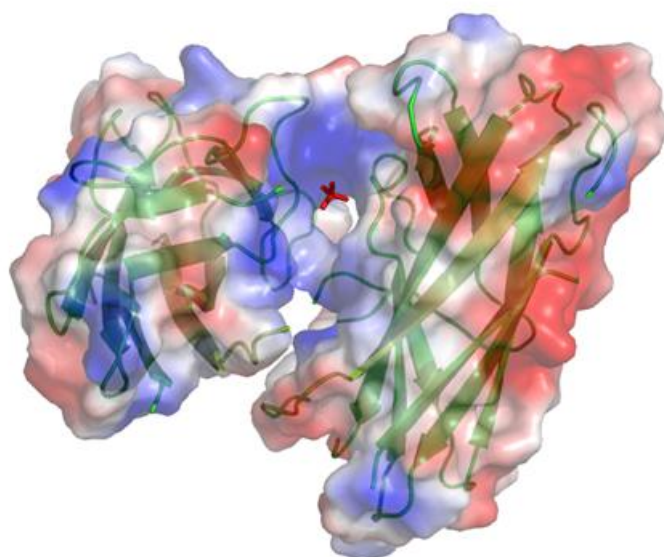


Figure 5.9: Crystals contacts between asymmetric units in the AafB unit cell. Different views of the contact between asymmetric units with different colours for each asymmetric unit.

5.6.3 Potential Binding Sites

An electrostatic analysis of the AafB structure reveals many negative and positively charged regions, especially on contact areas between the two molecules in the asymmetric unit (Fig. 5.10). Interestingly, a sulphate ion was observed in the centre of the positively charged interface between the two protein chains. The interaction observed might indicate a potential binding site. Other studies have suggested that ions and ligands bound within the crystal structure can be interpreted as highlighting a putative ligand binding sites (Kumar *et al.* 2010).

A)



B)

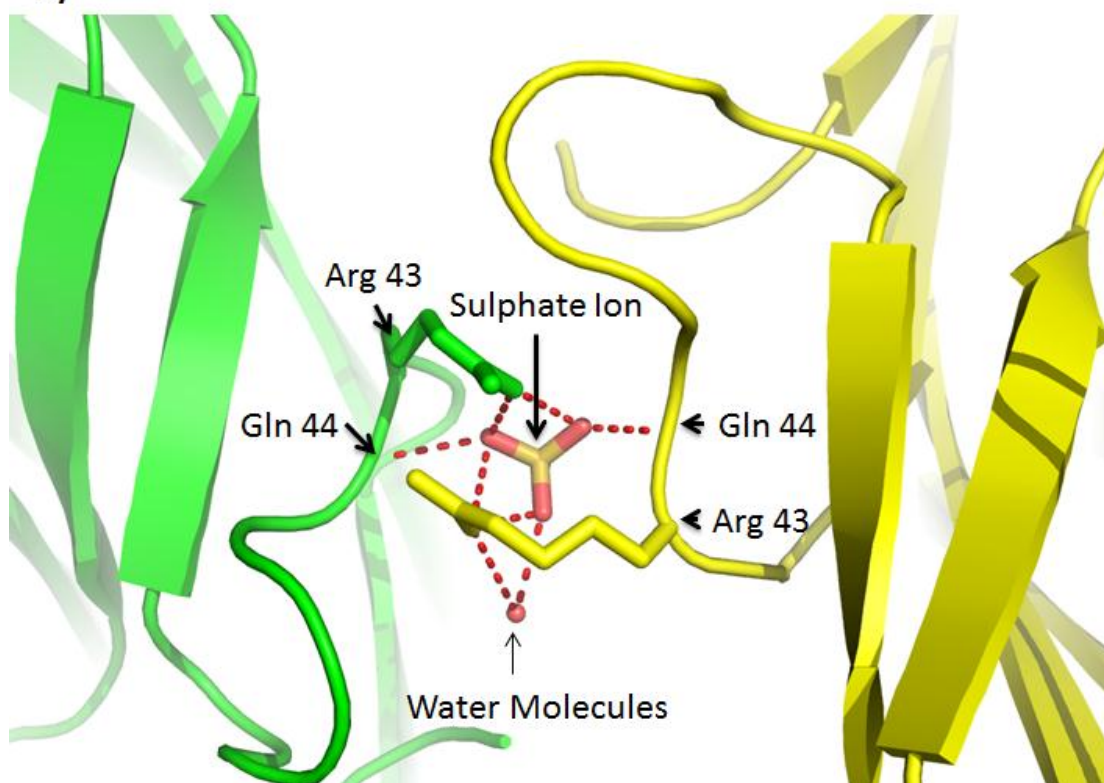


Figure 5.10: The sulphate ion contact with AafB. A) The electrostatic surface charge of AafB, the negative charge and positively charged areas are in red and blue, respectively. Beneath the surface there is a ribbon representation of the structure of AafB in green and the donor strand is in yellow. The sulphate ion is represented in stick with the colour red. B) The contacts between sulphate ions, water molecules and AafB are in red dash lines. The asymmetric units of AafB are coloured in green (chain A) and yellow (chain B). The oxygen atoms on the sulphate ion are coloured in red.

5.7 Comparisons of AafB with Afa/Dr Homologues

5.7.1 Sequence Alignment

Two known structures from the Afa/Dr family were used for the comparisons with AafB: AfaD and DraD from DAEC and UPEC, respectively (Fig. 5.11). The sequence alignment shows that AfaD and DraD are very similar, with the only differences being at the engineered donor strand which is used to stabilise the proteins. The protein sequence of AafB shows high similarities (57% identities) with many conserved residues, with disagreements mainly occurring at the designed donor strand (which is from the major subunit, AafA). The minor differences between AafB, AfaD and DraD were explored further by secondary and tertiary structure comparison.

```
AfaD      AELHLE-SRGGSGTQLRDGAKVATGRIICREAHHTGFHVWMNERQVDGRAERYVVQ-SKDG 58
DraD      AELHLE-SRGGSGTQLRDGAKVATGRIICREAHHTGFHVWMNERQVDGRAERYVVQ-SKDG 58
AafB      TEISLEGLHRNMGEQLFDGILATGRIICREAHHTGFHQMNAQVEGRPGHYIVQGSKDT 60
          *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

AfaD      RHELVRTGGDGSVPVKGEGGKGVSRPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVV 118
DraD      RHELVRTGGDGSVPVKGEGGKGVSRPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVV 118
AafB      QSKLVRLGREGWTSPVTGGGQGIIVRSQEEQVIFDVMADGNQWAKPGGEYIFSVSGKCLT 120
          *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

AfaD      PQEDNKQGFTPSGTTGTTKLTVT 141
DraD      PQEKLAAALEHHHHHH----- 134
AafB      SWEDNKQNATAVAKTATSTITVV 143
          *
```

Figure 5.11: Sequence Alignment of AafB with Afa/Dr family. Strong similarities were demonstrated by AafB with AfaD and DraD from the Afa/Dr family. The amino acids are coloured differently for easy observation of any disagreement between the sequences and the star under the residues indicated the present of identical amino acid at the position.

5.7.2 Secondary Structure Comparisons

Both AfaD and DraD display β -strands in their secondary structure with 12 β -strands, whilst AafB has only 9 β -strands. However, by comparing the regions of these β -strands, the positions and the residues are very similar amongst these three proteins. In region 2 and 7 the extra β -strand from AfaD likely corresponds to the longer β -strand present in both AafB and DraD (Fig. 5.12) which would be confirmed by

tertiary structure comparisons. Noticeably some β -strands from AfaD and DraD were not found in the equivalent position in AafB, instead loops were observed in AafB. In addition, the 9th β -strand of AafB was the engineered donor strand therefore differences would be expected as the sequence alignment shows many differences.

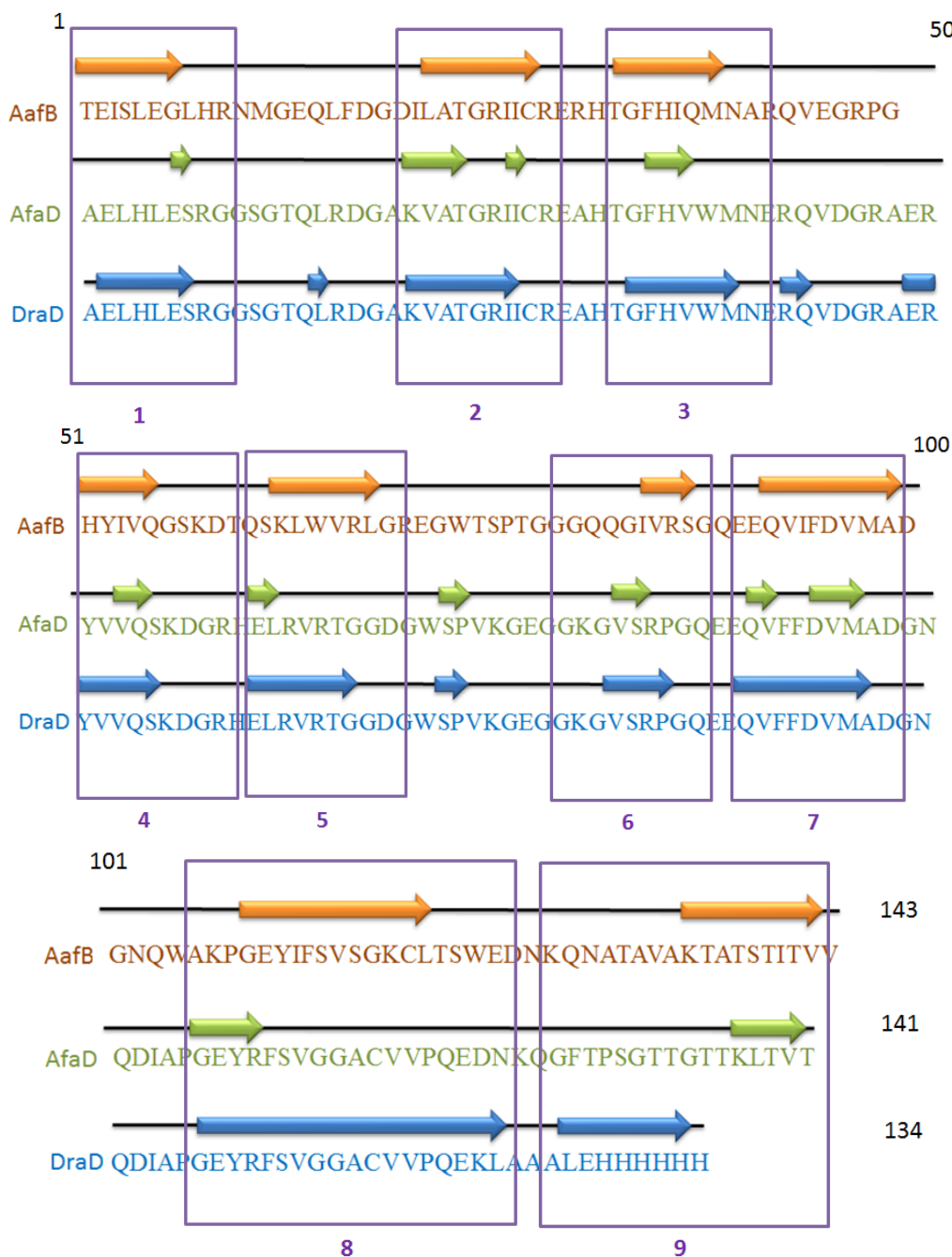


Figure 5.12: Comparisons of Secondary Structure of AafB with Afa/Dr Family. The arrows represented β -strand and the purple areas with numbers are the approximate locations of the 9 β -strands from AafB. AafB is in orange, AfaD is in green and DraD is in blue.

5.7.3 Tertiary Structure Comparisons

The structure of AafB was superimposed with AfaD (PDB: 2FVN) and DraD (PDB: 2AXW) and shows a high degree of similarity in the backbone with rmsd of 1.891 Å and 0.635 Å, respectively (Fig. 5.13). However, disagreements are observed in most of the loop regions.

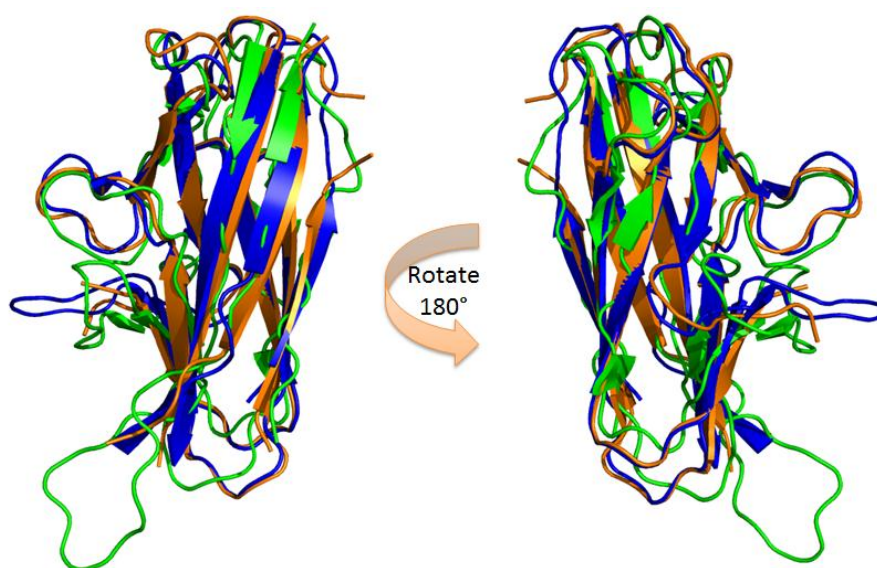


Figure 5.13: Superimposition of AafB with AfaD and DraD. AfaB is coloured in orange, AfaD is in green and DraD is in blue. Overall structures were similar with some differences in the loop regions.

Separate comparisons between AafB with AfaD and AafB with DraD were also carried out (Fig. 5.14). AafB is very similar to DraD as almost all the β -strands aligned perfectly and some loops were either at the identical position or followed the same orientation and directions. The greatest disagreement between the DraD and AafB occurred on the locations near the donor strand region and areas around the missing residues in the AafB structure. The three additional β -strands on DraD are at the position of loop regions on AafB (L1, L2 and L3 Fig 5.6), and as these extra β -strands were short and the structure of AafB is not at high resolution the contacts between β -strands might not be spotted hence loops were modelled to those areas

instead. As with AafB and AfaD, despite the overall fold being very similar, many loop regions are very different; especially around the long loops where these positions and shapes are very diverse. Many short β -strands are seen in AfaD and some of these were superimposed on the longer stretch of β -strands in AafB. In addition, the native sequence of AfaD and DraD were identical, therefore one of the reasons for the big differences in AafB and AfaD could be the different techniques (NMR and X-ray crystallography) used for determining the structures.

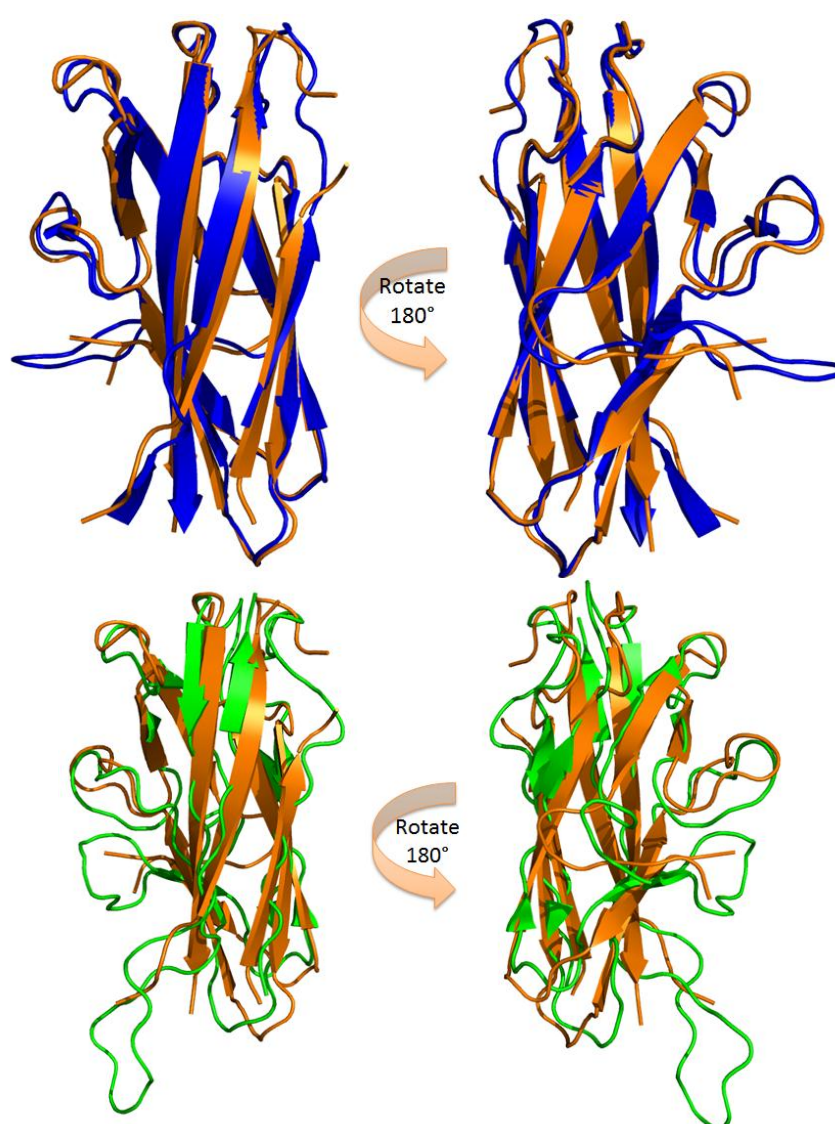


Figure 5.14: Comparisons of AafB with AfaD and DraD. AafB and DraD showed high degree of similarity in tertiary structure as most of the β -strands and some loops were located at the same position. AafB and AfaD shown similar fold but most of the loops were more diverse. AafB is coloured in orange, DraD is in blue and AfaD is in green.

5.8 Summary

The structure of AafB was solved by X-ray crystallography to a resolution of 3 Å, it consists of 9 β -strands. The space group is P6₁22 and the final $R_{\text{work}}/R_{\text{free}}$ is 0.22149/0.24840. All the residues fall within the favoured or allowed regions of Ramachandran plot. Two molecules were found in the asymmetric unit which are very similar with some minor differences around the missing residues which were in acceptable as they were the loop regions adjacent to residues with high B-factors.

Two ions were found (sulphate and acetate) interacting with AafB which might facilitate the overall crystal packing, but more importantly these protein regions are positively charged hence could potentially function as binding sites. However, there has been no binding site reported yet for AafB or the equivalent proteins (fimbrial minor subunit) therefore further studies are needed to confirm this.

The comparisons of AafB with AfaD and DraD from the Afa/Dr family confirm the presence of an immunoglobulin-like domain. Including the DSC strand, AfaD and DraD contains more β -strands (12) than AafB (9) but as the comparison showed, some of the extra short β -strands from AfaD belong to the longer β -strands in AafB. However, in the comparison of AafB with DraD indicates, DraD has three additional β strands. In addition, most of the differences between the structures were at the loop regions.

Chapter 6 Functional studies of AafB

With the knowledge of the AafB structure obtained in Chapter 5, functional studies were pursued. As part of the AafB preliminary screening studies, an interaction of AafA/B subunits with various oligosaccharides was observed (see below). NMR was subsequently used in an attempt to elucidate the interaction further and this is the main focus of this chapter.

6.1 Interaction Screens

To investigate any potential carbohydrate interactions, protein samples of AafB, AafA (provided by Dr Yi Yang, Imperial College) and AafBA were submitted to our collaborators for oligosaccharide screening. The initial screen was performed with AafBA since it contained both AafB and AafA monomers so anything binding with either domain within the dimer would be detected. Amongst several families of carbohydrates exhibiting a positive hit, the group of glycosaminoglycans (GAGs) was identified as binding with the highest affinity (Fig. 6.1) and would thus be the first target of interaction studies.

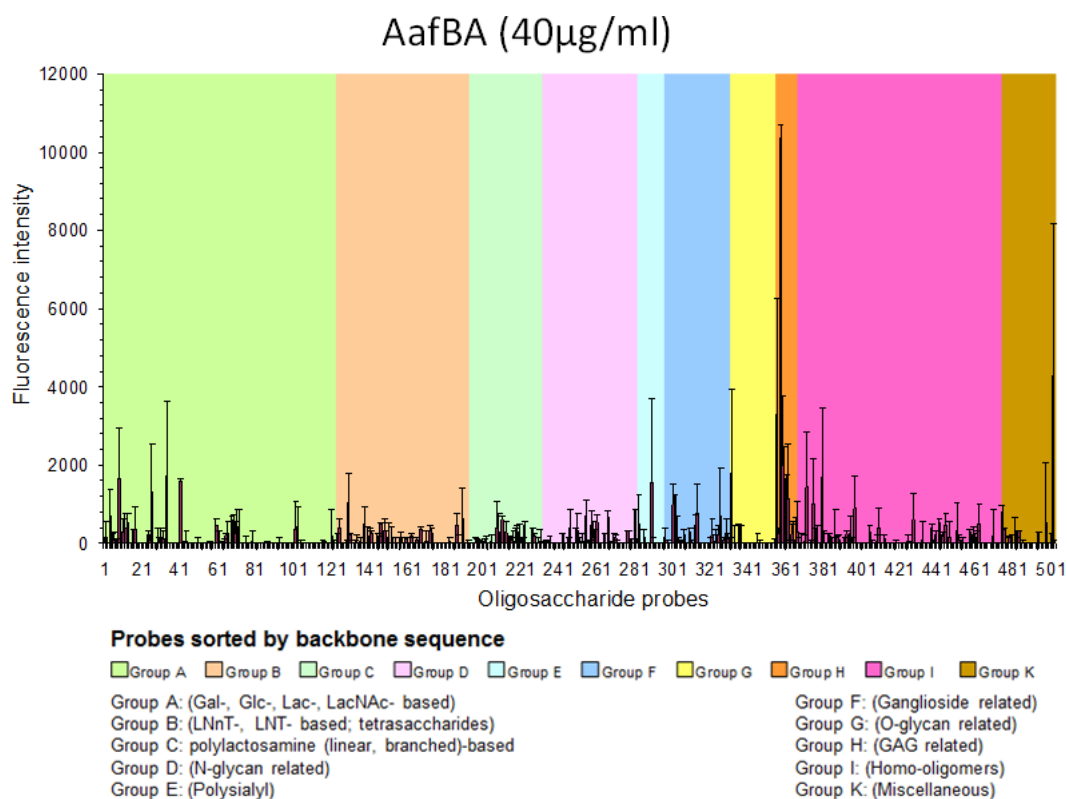


Figure 6.1: Microarray oligosaccharide screens of AafBA. Higher fluorescence intensity indicates more binding. Related saccharides are grouped and indicated by colour, shown below.

After this group was identified, different polymers of various disaccharides were used for an additional screen with all three Aaf constructs. These included chondroitin sulphate A (CSA), chondroitin sulphate B (CSB), chondroitin sulphate C (CSC), hyaluronic acid (HA), heparin sulphate (HS) and heparin (Fig. 6.2). No binding was detected for AafA with any of the GAGs, whilst AafB showed some evidence of a weak interaction with heparin and chondroitin. However, AafBA displayed much stronger binding with the GAG group, especially with heparin and chondroitin. To validate these preliminary findings and elucidate the putative carbohydrate binding site, binding studies were carried out using NMR spectroscopy. As the AafAB construct was likely to be too large for routine assignment methods, it was decided to pursue assignment of each individual subunit to permit monitoring of the interaction in the context of the 2-domain construct.

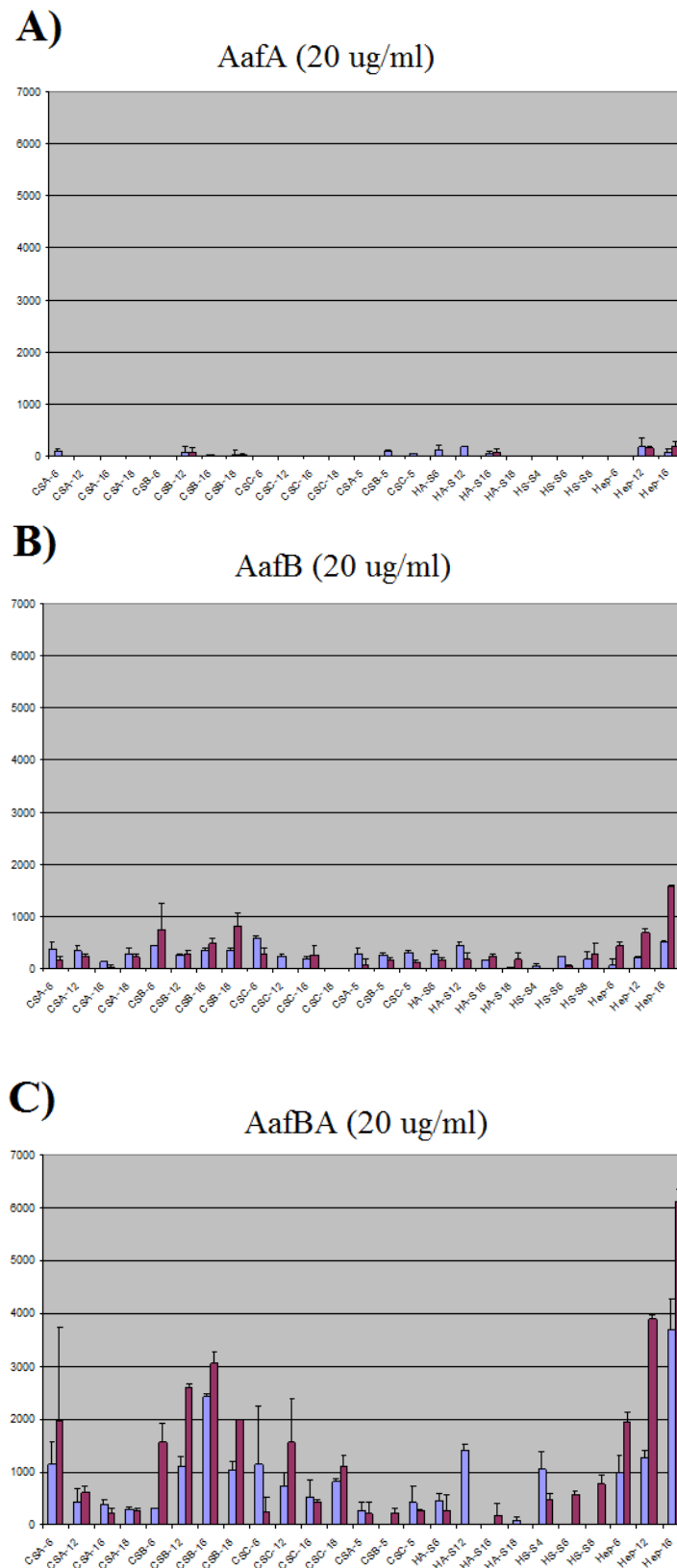


Figure 6.2: Microarray screens of Aaf with the GAG group of oligosaccharides. Different polymers were used and the lengths of these are indicated by the number on the label (blocking: casein (1/50) + BSA 1%). A) No binding was detected for AafA with any of the GAGs. B) Weak binding was detected with some heparins and CSB for AafB. C) Strong evidence of binding especially against heparin and CSB was detected for the 2-domain construct.

6.1.1 Glycosaminoglycan (GAG)

The glycosaminoglycans (GAGs) are long unbranched polysaccharides consisting of a repeating disaccharide unit which is made up of a six-carbon sugar (hexose or hexuronic acid) linked to another six-carbon sugar with nitrogen (hexosamine). The GAG group is an important component of connective tissues, one of the most common types of tissue found in the body. Both heparins and chondroitins are members of this family and can be found in joints in the body.

6.2 Nuclear Magnetic Resonance Studies

6.2.1 Suitability for NMR

A protein sample of AafB (with DSC strand) was produced identically to the samples used in structure determination in Chapter 5 and the protein buffer conditions described in Chapter 3. A 1d ^1H NMR revealed well-dispersed signals at 298 K with chemical shifts around the methyl and amide regions supporting a folded structure and also the presence of beta structure (as expected (Fig. 6.3a)). To test the stability of AafB under these conditions, the sample was incubated for up to two weeks at 37 °C with no signs of precipitation. A similar 1D NMR spectrum was observed (Fig. 6.3b). This indicated conditions suitable for longer 3D NMR experiments and also the potential of performing experiments at higher temperature to improve the signal to noise ratio if needed.

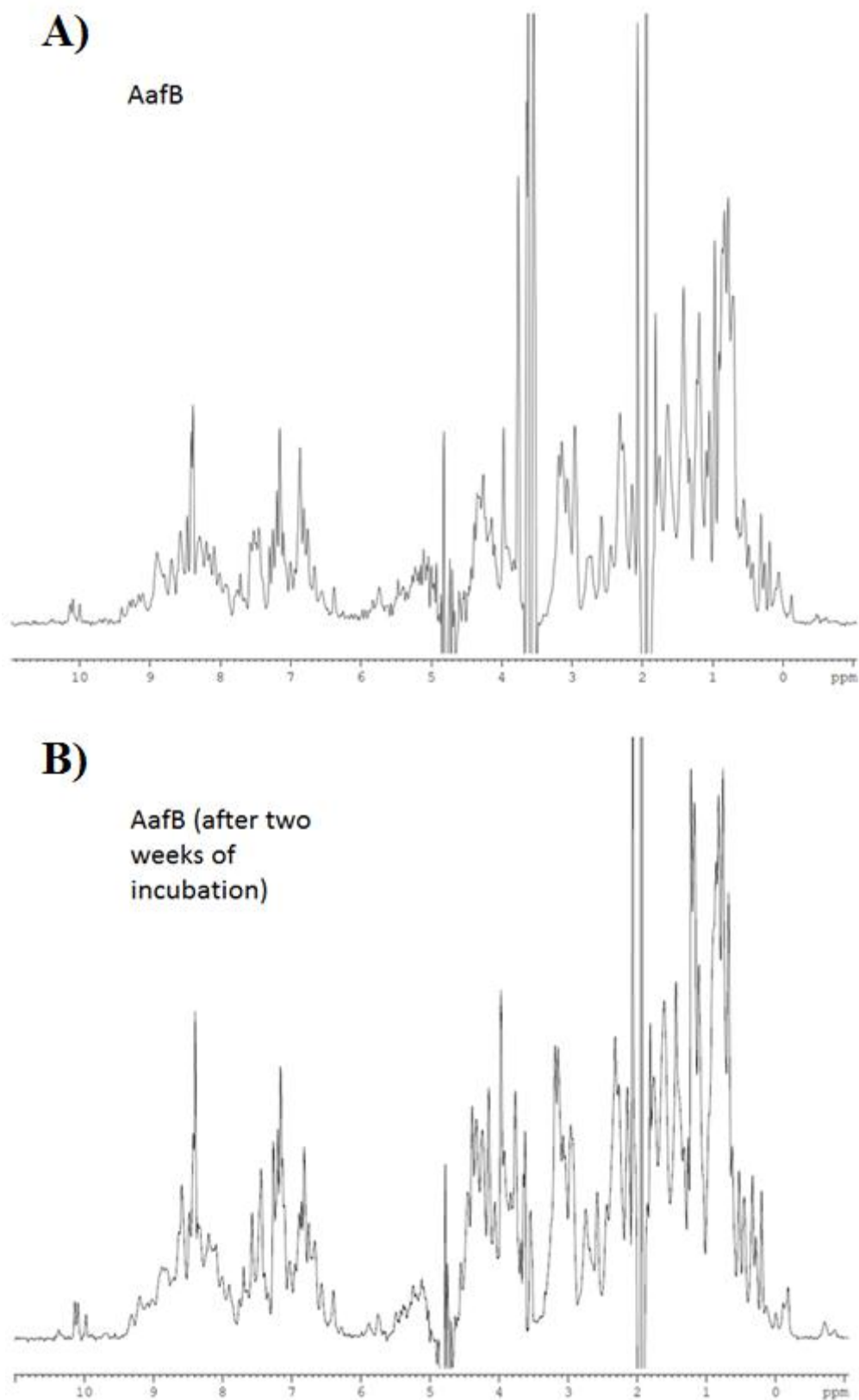


Figure 6.3: 1D spectra of AafB before and after two weeks of incubation. a) One-dimensional NMR spectrum of AafB which shows a folded protein as ring-current shifted peaks can be observed around 0 ppm due to structured methyls/methylenes and dispersed amides above 9 ppm. b) After incubation at 37 °C for two weeks, the spectrum remains essentially identical.

6.2.2 2D NMR

To investigate the feasibility of the 3d NMR experiments required for assignment and isotope labelling, approximately 5 mg of sample was produced from 1 litre of minimal media with ^{15}N ammonium chloride. A 2d ^1H - ^{15}N HSQC experiment was performed at 298 K (method in chapter 3) which revealed a very promising spectrum with well dispersed peaks (Fig. 6.4) indicative of a folded domain. The line width also confirmed a mono-disperse sample. Based on the AafB sequence and the flexible his-tag at the N-terminus, around 207 peaks were expected in the 2d ^1H - ^{15}N HSQC spectrum. However, over 250 peaks were found, which might be indicative of multiple species being present or conformational heterogeneity.

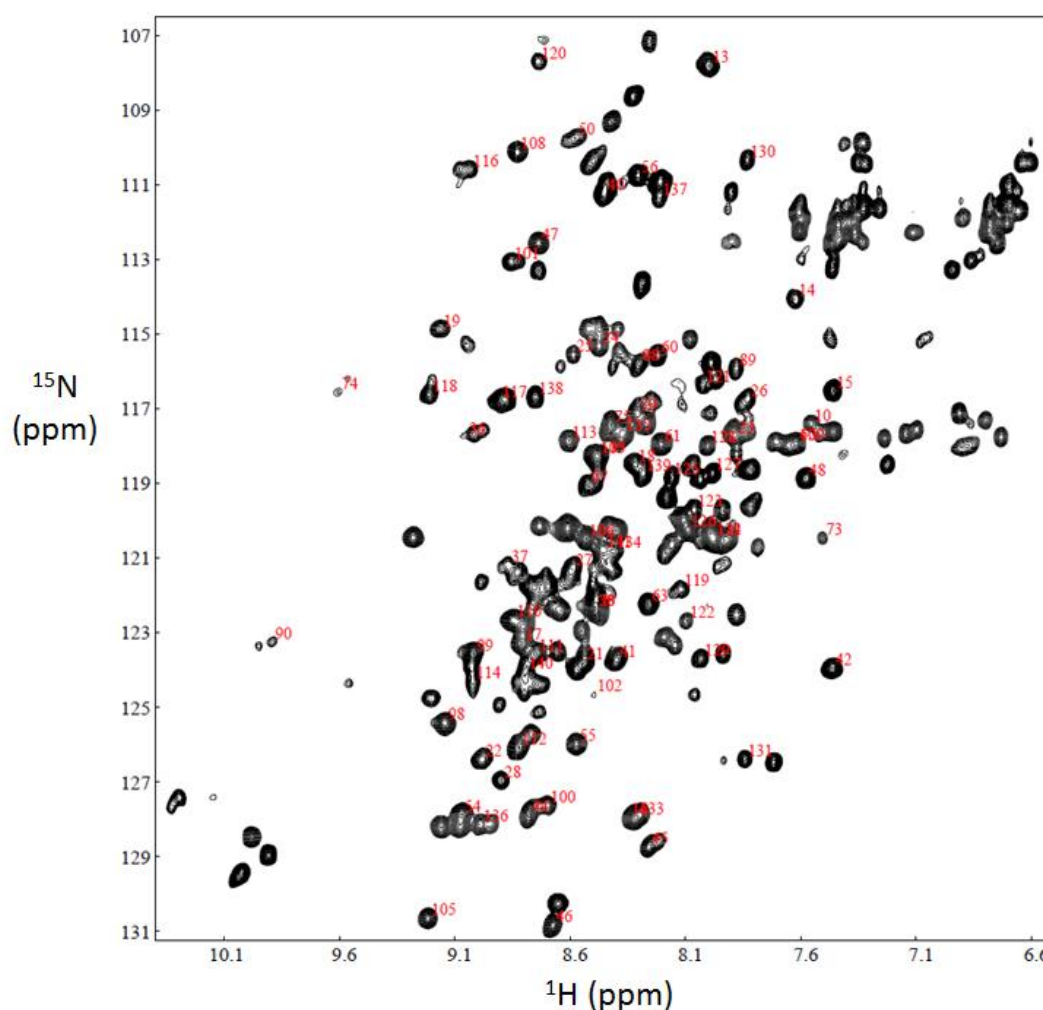


Figure 6.4: ^1H ^{15}N HSQC of AafB. Assigned peaks are marked in red with residue number.

6.2.3 NMR Backbone Assignments

After expression of a ^{15}N , ^{13}C labelled sample of AafB (described in Chapter 3) assignment was attempted by the methodology described by Sattler *et al* (Sattler 1999). For sequential backbone assignment, four 3-D NMR experiments were performed: CBCA(CO)NH, HNCACB, HNCO and HN(CA)CO and for reference a high-resolution ^{15}N ^1H HSQC was also recorded. Since 3-D experiments are less sensitive than 2-D experiments and in order to gain optimal signal to noise, a series of HSQC experiments were measured with different temperatures before carrying out any 3-D NMR work. In general, higher temperatures lead to a higher signal to noise ratio due to increased thermal motion and hence tumbling of the macromolecule, although specific proteins may not be stable at elevated temperatures or may be shifted to alternative conformations. Temperatures of 298 K, 303 K, 308 K and 313 K were tested and the spectra were identical, albeit with some minor shifting of peaks (Fig. 6.5). As the 3-D assignments were performed for determining the involved residues involved in the interaction, to prevent any unforeseen issues 308 K was chosen to perform all further 3-D experiments.

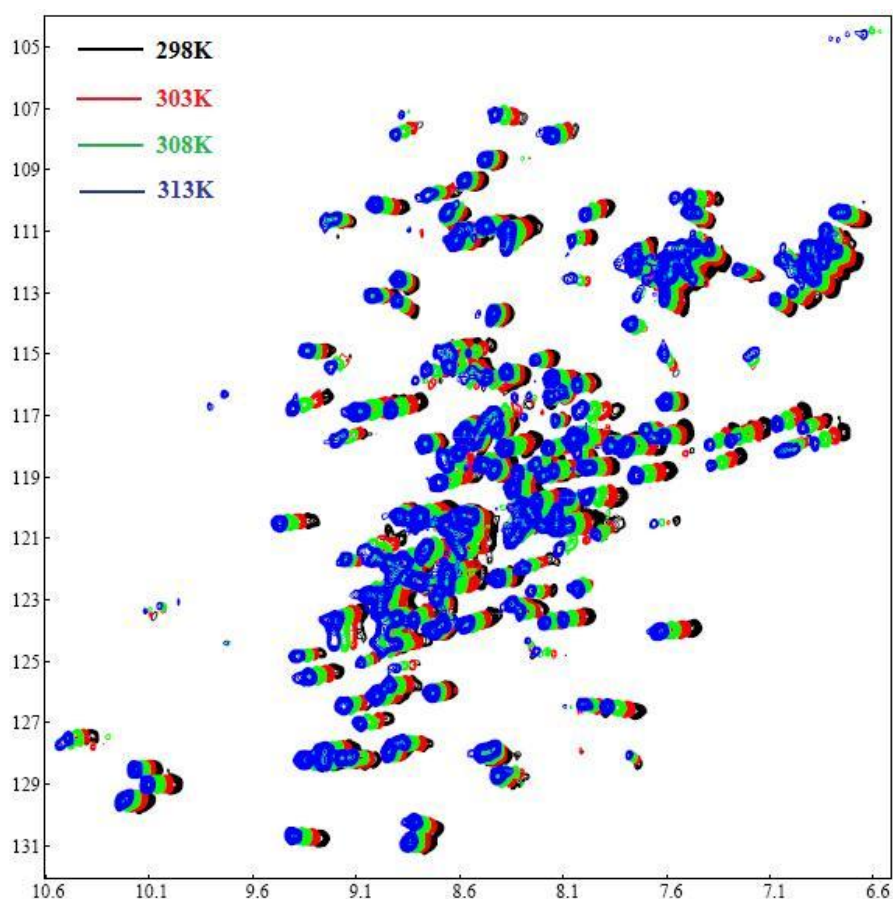


Figure 6.5: Overlay of HSQC spectra of AafB at different temperatures. 298 K (black), 303 K (red), 308 K (green) and 313 K (blue) show no changes in the number of peaks or any sign of unfolding or thermal denaturation.

6.2.3.1 HNCACB and CBCA(CO)NH

The CBCA(CO)NH and HNCACB can be displayed in “strips” according to the chemical shift from each amide proton detected in the HSQC. In the strips of the HNCACB, $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ resonances from a particular residue (i) and preceding residue (i-1) would be observed. To distinguish i from i-1 resonances we can see the strips from the CBCA(CO)NH which contain peaks only for i-1. Example strips are shown for AafB, illustrating linking of the spin system for residue Val97 to Gly101 (Fig. 6.6). In theory using a combination of these two types of spectra one can assign most of the backbone atoms, but in reality, due to overlap and signal-to-noise

problems the HN(CA)CO and HNCO (see below) are also be needed to completely resolve all the backbone assignments.

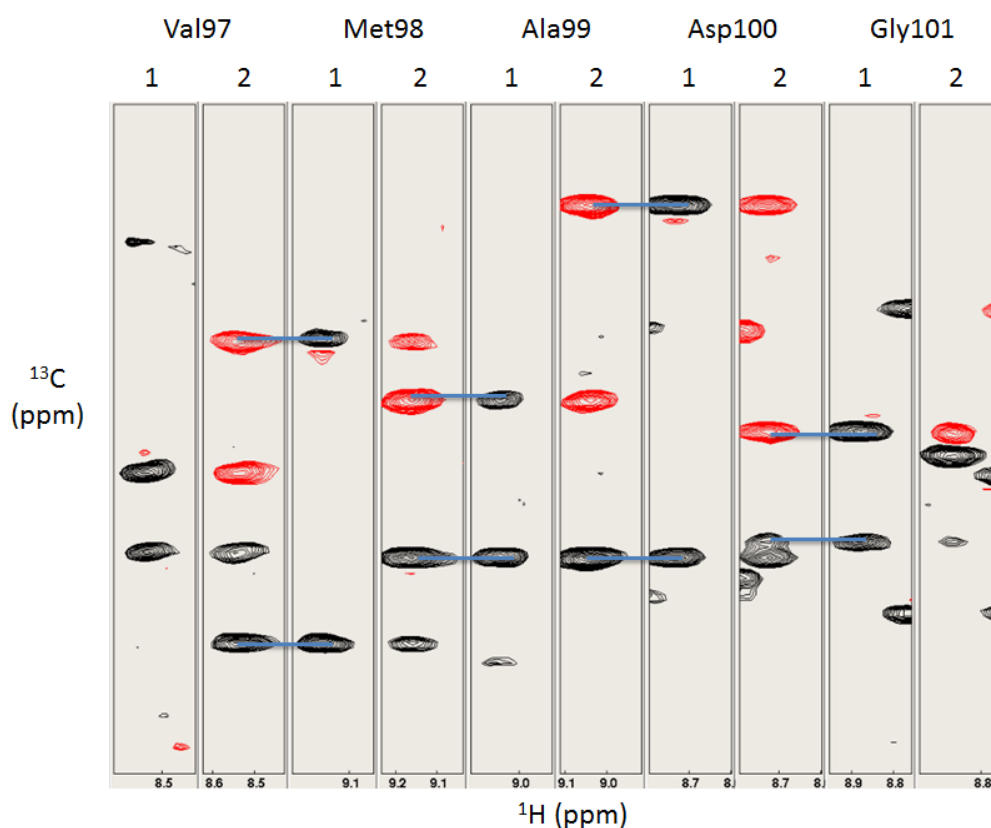


Figure 6.6: Examples of the HNCACB and CBCA(CO)NH strips. Strip 1 are from the CBCA(CO)NH spectrum and strip 2 are from HNCACB spectrum. $\text{C}\alpha$ and $\text{C}\beta$ peaks are coloured black and red, respectively. Blue lines indicate the connections between the residues.

6.2.3.2 HN(CA)CO and HNCO

In both HNCO and HN(CA)CO spectra only isotopically labelled carbonyl ^{13}C are detected. The HN(CA)CO displays both the ^{13}C chemical shift of a particular residue (i) and the residue before (i-1), and HNCO only displays the ^{13}C chemical shift of residue (i-1) (Fig. 6.7).

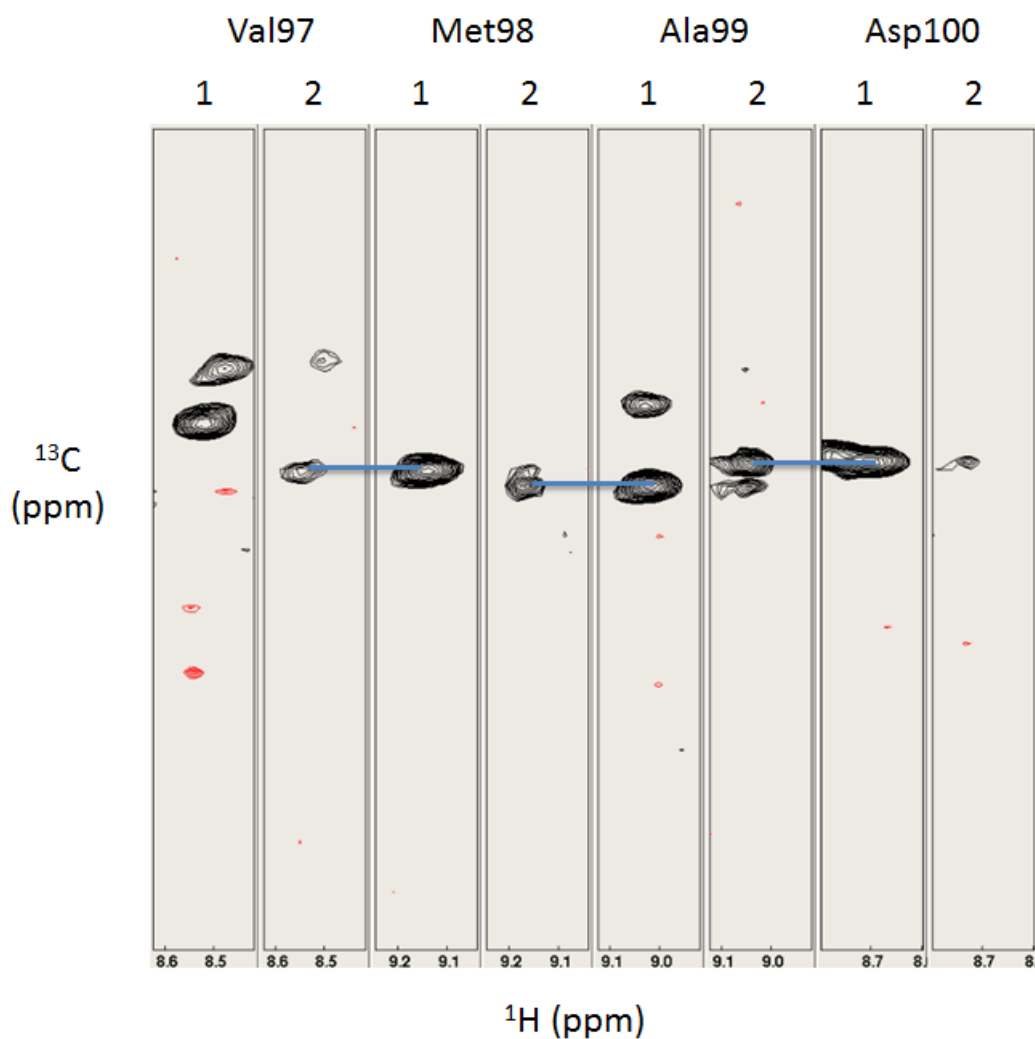


Figure 6.7: Examples of HN(CA)CO and HNCOS strips. Examples of identifying C^1 of a particular residue (i) and the residue before (i-1) by combining the strips from HN(CA)CO (strip2) and HNCOS (strip 1). Blue lines indicate the connections between the residues.

6.2.3.3 MARS Automatic Assignment

The initial backbone assignment was attempted using MARS automated assignment (Jung and Zweckstetter 2004). Excluding prolines MARS only assigned 45% of residues with high confidence, therefore further manual analysis and assignment of the strips were performed using NMRview. Several rounds of manual assignment and MARS were performed which ultimately lead to assignment of around 65% of the residues.

Manual analysis reveals significant problems arising due to extra “duplicated” peaks which increased overlap and ambiguity of assignments. The issue is potentially caused by conformational exchange or the presence of locally misfolded protein (see below).

6.2.3.4 Multiple Species

The backbone experiments were more complex than expected as many more peaks were observed than expected and this might due to the effects of conformational exchange manifesting itself as duplets of many peaks (Fig. 6.8). However, conformational exchange seems unlikely as altering temperature (ranging from 298 K-313 K) did not reduce the “doubling” effect (Fig. 6.5). However, the prolines might be at different state (either cis or trans) as the results of local conformational change. This suggested a possibility of multiple species in the sample, which shared very similar overall structure but also exhibit some minor differences in some regions. With around 65% assignments, two-thirds of the residues are covered and might be sufficient to continue interaction studies by NMR.

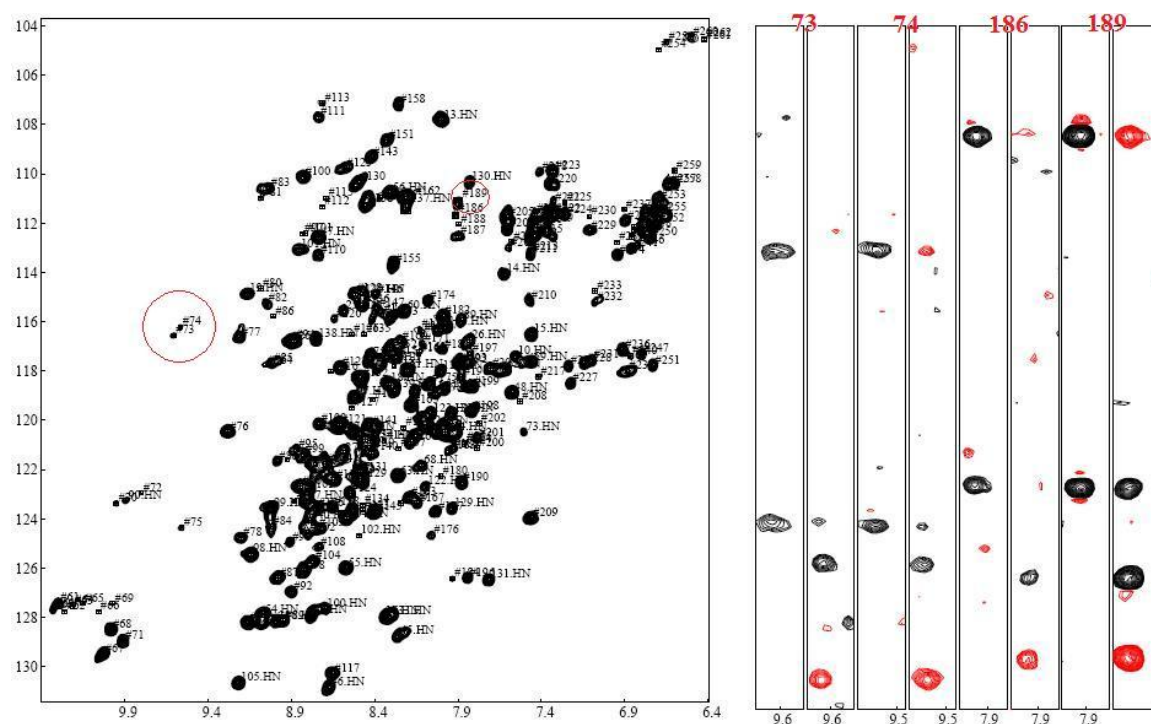


Figure 6.8: Example of doubling of residues in spectra of AafB. On the left is the HSQC. The circled peaks illustrate examples of doubling effect. On the right is the 3D analysis, confirming the doubling effect, as peak 73, 74 and 186, 189 have the same pattern in the C α and C β resonances.

6.2.4 Chemical Shift Index Analysis

With the limited assignment available a chemical shift index (CSI) analysis was performed for the secondary structure of AafB to confirm the solution state agrees with that observed in the crystal structure (Fig. 6.9). The data are in general agreement except for the α helical region which was designed as a loop linking the native protein with donor strand. In general, the chemical shift index analysis displayed multiple β strands as expected but for presumably the same reason the length of each strand was not always reliable in this case.

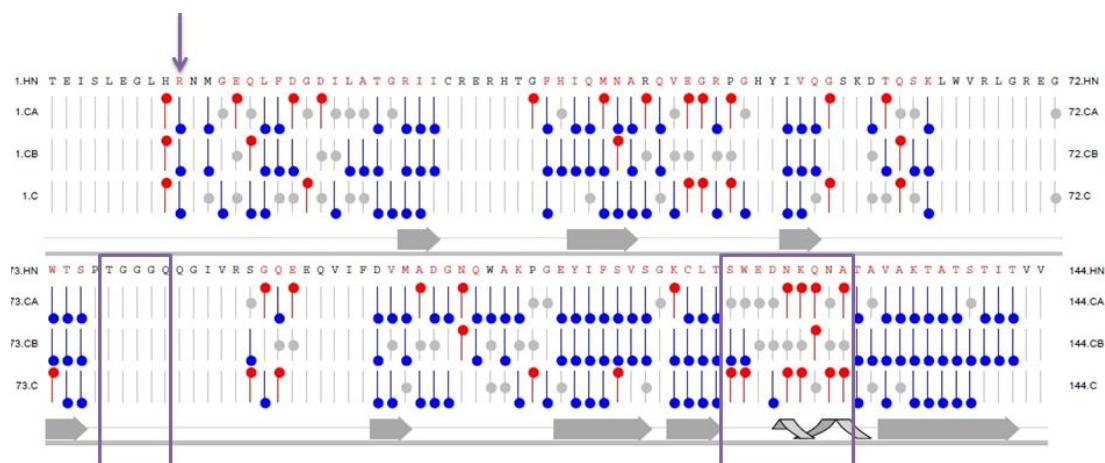


Figure 6.9: Secondary structure analyses using the Chemical Shift Index for AafB. The purple arrow and squares indicate the missing loop residues from the AafB crystal structure. The amino acids in red are the assigned residues and the unassigned residues are in black. The grey arrows and spiral indicate β strand α helix, respectively.

6.2.5 Mapping Missing Regions on AafB Structure

To investigate the cause of doubling effect, the missing residues were mapped onto the AafB crystal structure (Fig. 6.10). However, the missing residues were spread across the structure including several loops and β strands. It is likely the percentage of unassigned residues is higher than the percentage of “doubled” residues due to difficulties in assigning these residues hence it is still unclear what the causes the doubling effect. Furthermore, most of the assigned residues show similar secondary structure in CSI and the crystal structure (Fig. 6.11).

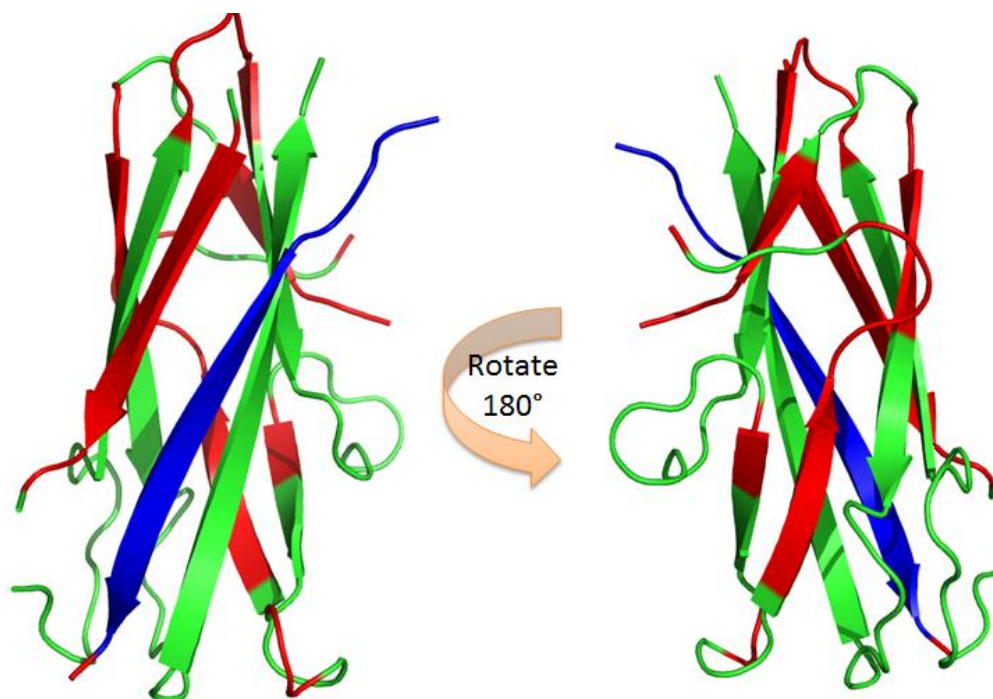


Figure 6.10: Missing Residues from NMR assignment mapped onto the AafB crystal structure. The residues that were assigned by NMR are coloured green, the donor strand is in blue and the missing residues are in red.



Figure 6.11: Comparison of experimental secondary structures of AafB. The assigned residues are in red and the unassigned residues are in black. The secondary structure from X-ray crystallography data and CSI is in orange and blue, respectively. The arrows and wave indicate β strand and α helix, respectively.

6.3 Interaction Studies

6.3.1 ^1H NMR of AafBA

The AafBA construct was ^{15}N labelled, in a similar manner to the unlabelled sample of AafBA described in Chapter 4. The 1d ^1H NMR of AafBA revealed well-dispersed signals at 298 K with dispersion of chemical shifts typical of a folded domain and also the presence of β -structure. The spectrum was also overlaid with the AafB spectrum and both shared similar peaks around the methyl and amide regions, the differences being presumably due to the presence of the AafA domain. However, the line widths remained similar indicating the independent motion of the two domains.

6.3.2 Interaction of AafB with AafA

Despite difficulties in transferring assignments of AafB to the two domain construct due to some changes in chemical shift and extra signal overlap, it is possible to map out a few residues that were shifted when AafA is linked to AafB. Most peaks were unchanged, indicating the interaction did not significantly change the tertiary structure of AafB (Figure 6.12).

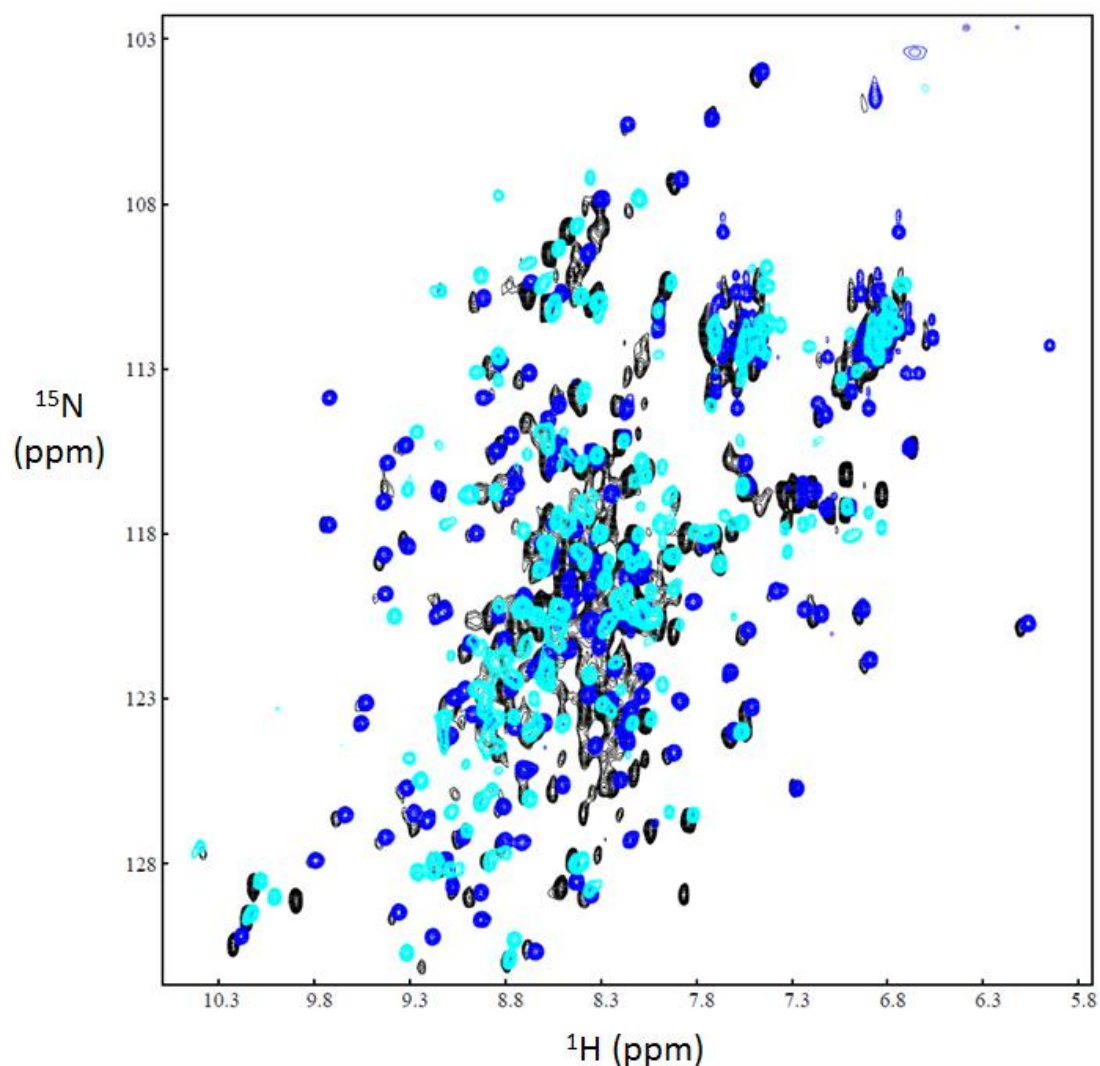


Figure 6.12: Overlaid HSQC spectra of AafBA with AafB and AafA. The HSQC of AafBA is in black. The HSQC of AafB and AafA are in cyan and blue, respectively.

Changes in chemical shifts can be identified in AafB when compared to AafBA and were mapped onto the AafB structure (Fig. 6.13). They mapped to the region of expected interface to AafA and near the donor strand. This was expected since the donor strand was designed to mimic the strand exchange mechanism between domains and the interaction was expected to allow the insertion of this strand into the domain. Furthermore, as AafB is found on the tip of the fimbriae with a polymer of AafA extended below it, the lower region would be close in space to AafA, hence some kind of interaction and change of chemical environment could be anticipated, leading to peak shifts. However, as the line-width remained similar in HSQC, there

may be a degree of independent motion of the two domains which could be caused artificially by the long non-native linker in the construct. In addition, despite some shifting of peaks, the other non-AafB peaks corresponded to the peaks of AafA (provided by Dr. Y. Yang, Imperial College) which confirmed the presence of the two domains in their native state.

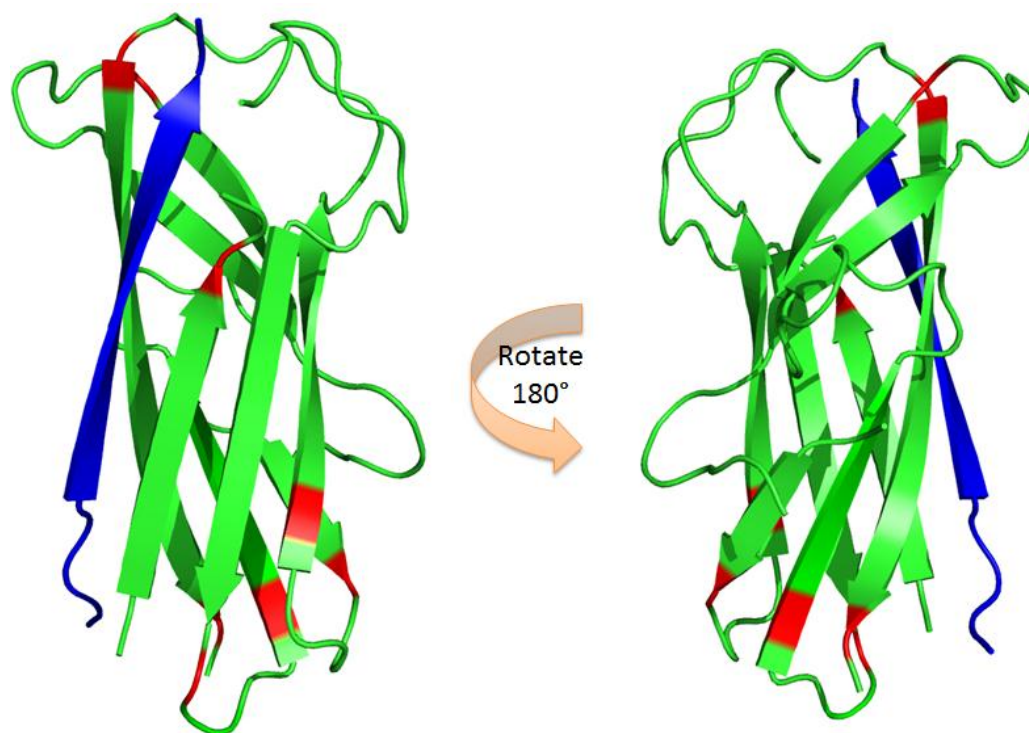


Figure 6.13: Residues of AafB affected by AafA. The structure of AafB with the donor strand in blue and the residues that were shifted due to the interaction with AafA are mapped in red.

6.3.3 Titration of AafBA with Chondroitins

The same buffer was used for both AafBA and all chondroitin derivatives and measurements confirmed that the pH remained unchanged after mixing. However, mixing AafB with any chondroitin derivatives lead to heavy precipitation. Therefore, AafB was dialysed into a range of buffers between pH 3 and pH 6 prior to mixing but still heavy precipitation occurred after adding any chondroitin derivatives. HSQC spectra of the supernatants were recorded and protein remaining in solution showed

no peak shifts, hence no interaction was detected (Fig. 6.14). However, as the solution conditions were not changed by adding chondroitins presumably this indicates that some sort of interaction led to precipitation and aggregation which was not detectable by NMR.

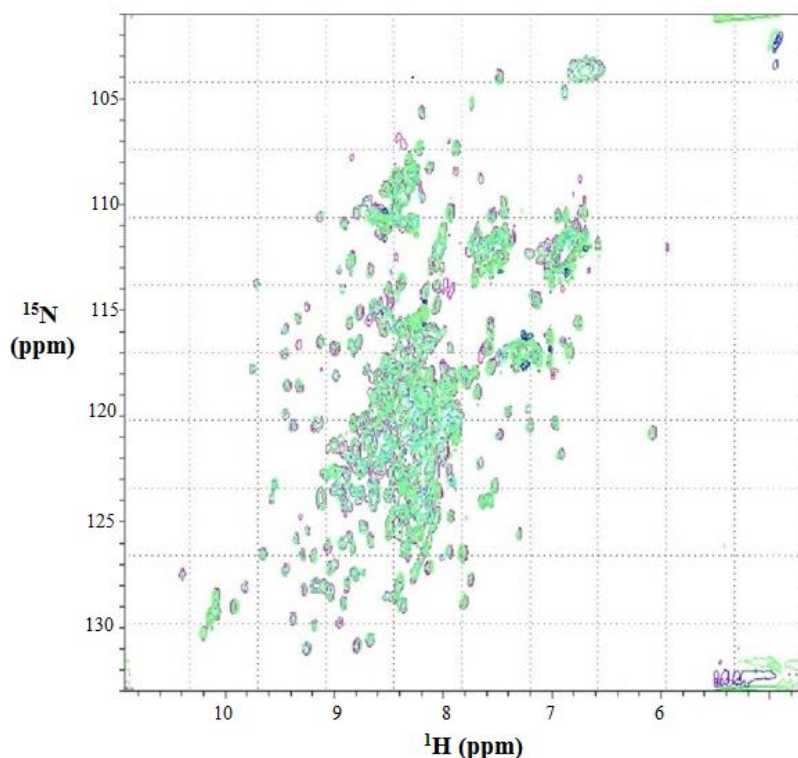


Figure 6.14: An example of titration of AafBA with chondroitin sulphate C (CSC). No peak shifting was observed in the protein that remained in solution. AafBA is coloured purple and the spectrum after adding CSC (1:5) is in green.

6.3.4 Titration of AafBA with Heparin

Different heparin derivatives are commercially available and were used in the following titration (Fig. 6.15). The buffer used to dissolve heparin derivatives was the same as used for AafBA, and the pH remained unchanged after dissolving heparin derivatives. Interestingly, mixing the AafBA with any heparin derivative did not result in precipitation. However, no peak shifting was observed from the titration of AafBA with heparin III-S and heparin I-A, hence there was no evidence of binding (Fig. 6.16).

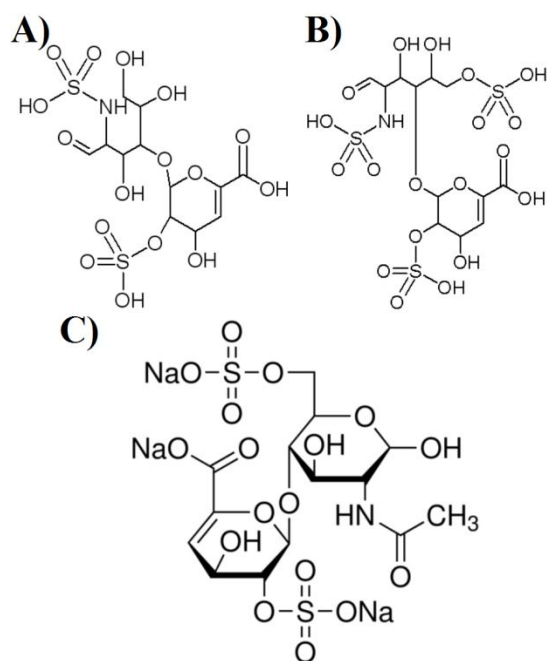


Figure 6.15: Structures of Heparin derivatives. Different Heparin derivatives used in titrations with AafBA; A) Heparin III-S, B) Heparin I-S, C) Heparin I-A.

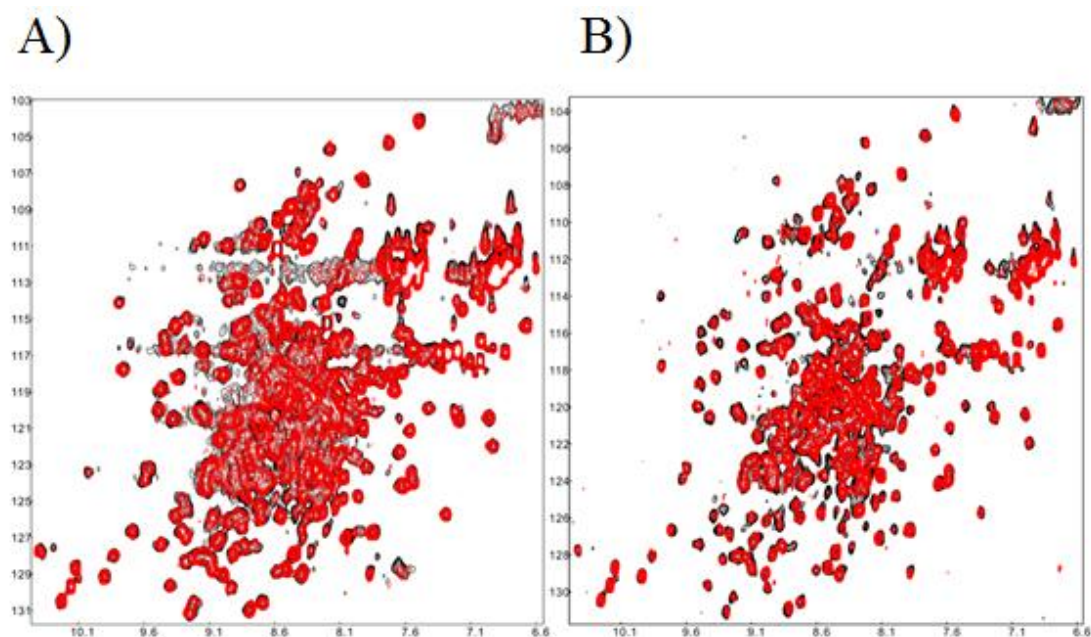


Figure 6.16: Titration of AafBA with Heparin III-S and Heparin I-A at 1:40 ratio. A) Titration of AafBA with Heparin III-S. B) Titration of AafBA with Heparin I-A. Both HSQC show no peak shifts, hence no interaction was detected.

A titration of AafBA with heparin I-S was also performed. In this case, only a handful of peaks shifted by a very small distance indicating at best a weak interaction (Fig. 6.17). With the incomplete assignment of AafB and increased peak overlap in AafBA, it was difficult to determine which domain (AafA or AafB) contained these peaks.

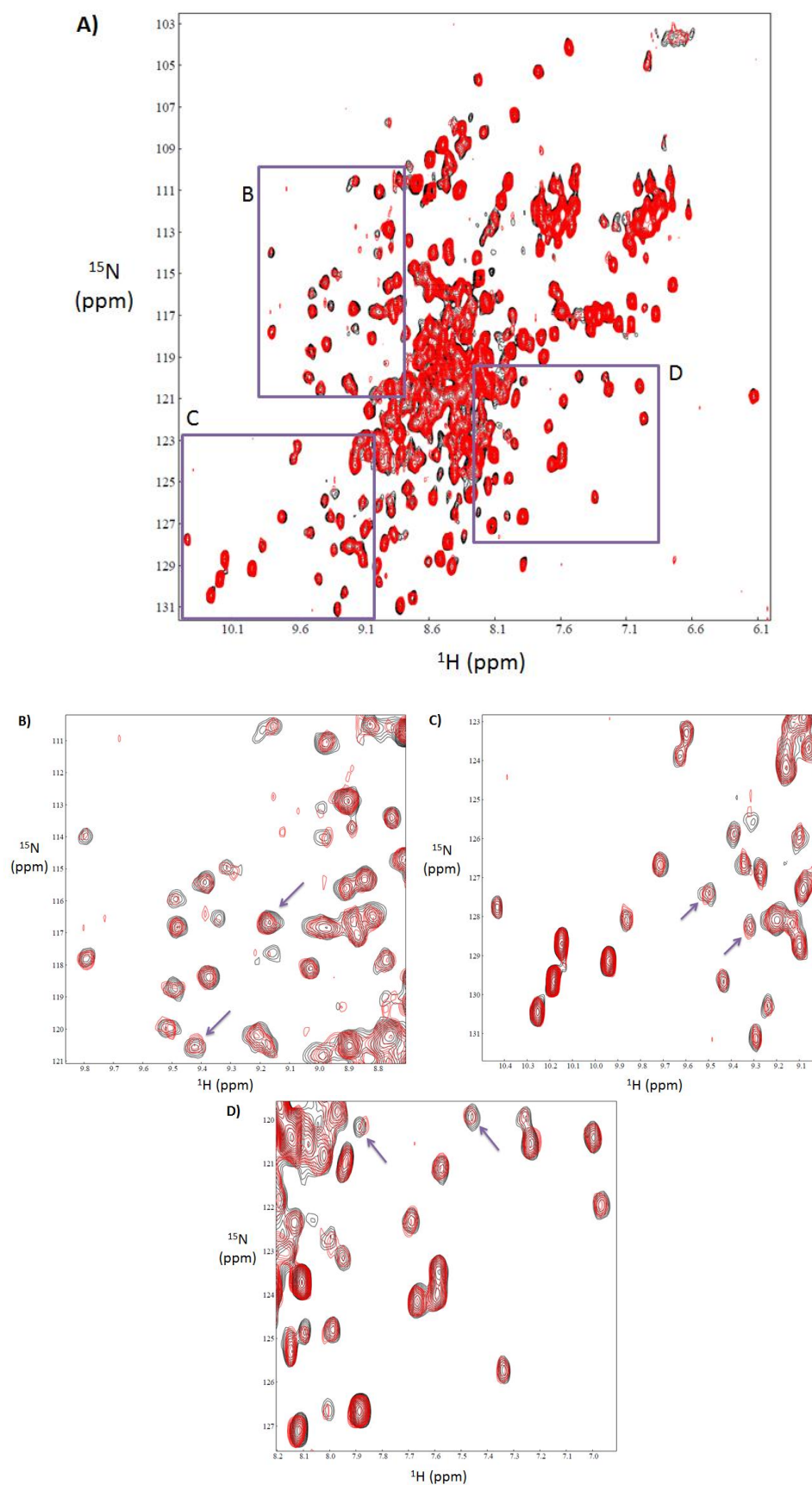


Figure 6.17: Titration of AafBA with Heparin I-S. A) AafBA is coloured black and the spectrum after adding heparin I-S (1:5) is in red. The peak shifting was very minor, some examples of which are highlighted in the boxed areas. B), C) and D) are the expansions of these areas where some shifting peaks are indicated by arrows.

6.3.5 AafB Interacting with Heparin I-S

Since the spectra of AafBA were complicated by the high degree of overlapping peaks, titration of AafB with heparin I-S was performed but no changes were observed (Fig. 6.18A). The missing peaks might be the result of strong exchange hence more likely to be involved in the interaction with heparin I-S, The other peak shifting could be the results of low exchange rate and less significant in the interaction. However, due to the limited assignment and overlapping peaks only three shifted peaks were identified, Gly 116, Thr 24 and His 37 (Fig. 6.18B). Histidine has a positively charged side-chain at pH 5.4 and since heparin I-S is highly negatively charged, His 37 could be involved in a direct electrostatic interaction with heparin I-S. Changes in the chemical shift of Thr 24 and Gly 116 could be the side effect of the changes in the binding especially Thr 24 is located on the opposite side of the molecule. However, AafB might be binding to more than one heparin I-S, hence multiple binding sites might occur. In addition, the residues (Arg 43 and Gln 44) involved in the interaction with sulphate ion from the crystal structure are not identified. This is due to Arg 43 was not assigned and Gln 44 was overlapping with AafA peaks, therefore unable to confirm the changes. However, both Arg 43 and Gln 44 are located near Gly 116 and His 37 which might assist binding or could possibly is an addition binding site.

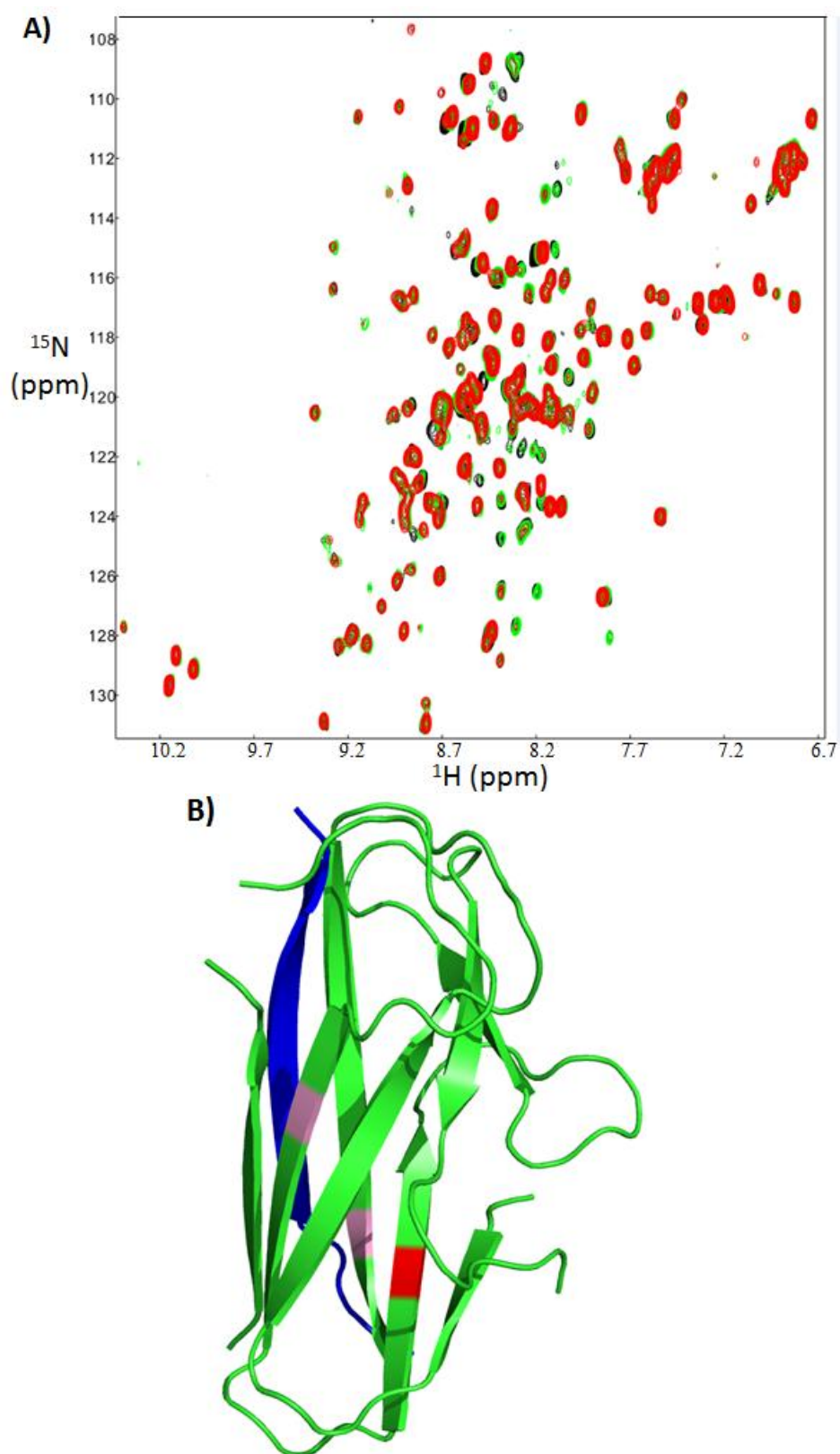


Figure 6.18: Titration of AafB with Heparin I-S. A) Overlay of AafB titration with Heparin I-S. The spectrum of AafB is in black and spectra acquired with different ratios of Heparin I-S are in green (1:10) and red (1:40). B) The structure of AafB where the donor strand is in blue and the residues identified in the titration of AafBA with Heparin I-S, threonine 24 and glycine 116 are in pink and histidine 39 is in red.

6.4 Summary

Microarray screens suggested several potential binding partners, in which chondroitins and heparin displayed the highest binding affinity. In order to confirm the interactions NMR experiments were performed.

The AafBA construct was expressed ^{15}N labelled but the HSQC spectra contained a high degree of overlap. Therefore, both AafB and AafA were also expressed and assigned separately. The structure of AafA was solved by NMR and the data was provided (Dr. Y. Yang, Imperial College) but due to the multiple species present in the AafB sample a “doubling” effect of the peaks was observed and lead to incomplete assignment (~65%). The HSQC spectra of AafBA, AafA and AafB were compared and confirmed the peaks on the AafBA spectra came from either AafA or AafB which indicated the presence of the two domains in a folded state. Furthermore, the line-width of AafBA was similar to AafA and AafB spectrum; this suggested the two domains might tumble independently in the context of the construct used, which contained an artificial linker.

Heavy precipitation was observed on mixing of AafBA with chondroitins and the condition did not improve by changing the pH of solution. The HSQC spectrum of the supernatant showed no changes when compared to the spectrum before addition of chondroitin. The precipitation indicates an interaction occurs but the nature of the interaction is unclear.

Different heparins were used for titration with AafBA. Addition of neither heparin I-A or heparin III-S resulted peak shifts, hence there was no evidence of binding under these conditions. Some minor peak shifts were observed on addition of heparin I-S to

AafBA, but the spectrum was heavily overlapped and was therefore difficult to determine which domain the affected residues were in.

With the limited assignment of AafB, only three residues were identified to be shifted in the titration of AafB with heparin I-S (Thr 24, Gly 116 and His 37). Interestingly, Gly 116 and His 37 are located near the interface where AafA would be linked to AafB. Hence, these two residues might be part of an intermolecular binding site of AafBA with the polymer of heparin, which displayed a much higher binding affinity in the microarray screening (6 to 12 polymers).

To confirm the hypothesis that these are the residues involved in the intermolecular interaction a polymer of heparin, further studies such as mutagenesis would be needed. As the microarray screens were performed with different length of polymer of heparin and chondroitins, this might suggest the binding required the presence of two domains. This could explain the negative results in the titration of individual domains.

Furthermore, co-crystallisation screens could also be performed with different polymers of heparin and chondroitins with AafB, as used in the microarray screens. The crystallisation of AafBA was unsuccessful which might due to the long linker between the domains as the NMR results suggested the two domains were likely to be in independent motion. However, by co-crystallisation with oligosaccharides such as heparin or chondroitins, the interaction might restrict the movement of the domains and facilitate the lattice formation.

Chapter 7: Structural Determination and Analysis of HdaB

7.1 Data Collection

Crystals of HdaB were soaked in cryoprotectant containing sodium iodide (the composition of cryoprotectant is given in section 3.4.3) for a few seconds then flash frozen in liquid nitrogen. These were taken to beamline I24 (Diamond Light Source, UK) and exposed to synchrotron radiation with wavelength 1.65310Å. The crystals were similar in size and diffracted to around 2.3-2.0 Å. After indexing 2 images separated by 90°, the point group was suggested to be P422 and in order to provide a highly redundant dataset for phase determination used in iodide-SAD, 720° of data were collected. The dataset was collected at 100 °K with an oscillation angle of 0.2° per frame to 2.0 Å on a Pilatus 6M detector, with a detector distance of 19.4 mm.

7.2 Data Processing

The data were first indexed in the point group P422 to 2.0 Å using MOSFLM (1994) and scaled with SCALA (1994). The results from SCALA were fed into Pointless (1994) which confirmed the point group of P422 and suggested space groups of P4₁2₁2 or P4₃2₁2 based on the strong systematic absences. Therefore, the data were reprocessed, scaled and truncated in space group of P4₁2₁2. The statistics for processing and scaling of the HdaB data are shown in Table 7.1. Analysis of the crystal composition gave a Matthews coefficient (V_m) of 2.86 and 1.91 indicating that there are two or three protein molecules in the asymmetric unit with a solvent concentration of 56.99% and 35.49%, respectively (Table 7.2).

Protein		HdaB (with HdaA DSC strand)
Space group		P4 ₃ 2 ₁ 2
Cell dimensions (Å)	a	112.2940
	b	112.2940
	c	61.6521
Cell angles (°)	α	90.00
	β	90.00
	γ	90.00
Number of reflections (total)		482265
Number of reflections (unique)		27204
I/sigI (mean)		5.4 (1.9)
Multiplicity		3.6 (3.1)
R _{sym} (%)		8.0 (37.1)
Completeness (%)		98.1 (95.1)
Resolution range (Å)		28.74 - 2.00 (2.11 – 2)
Average B Value (Å ²)		34.062
RMS Correlation Ratio (RCR) of anomalous signal		0.0966 (0.962)
Rpim		0.046 (0.237)

Table 7.1: Summary of data collection and scaling statistics for HdaB. Statistics for the outliers are in brackets.

Number of molecules per Asymmetric unit	Matthews Coefficient (Vm)	Percentage of Solvent Content
2	2.86	56.99
3	1.91	35.49

Table 7.2 Summary of Matthews Coefficient of HdaB. The number of molecules per asymmetric unit is given with the corresponding Matthews coefficient score and solvent content.

7.3 Structural Determination

Although an anomalous signal from iodide was measured, it was not sufficient for phase calculation and therefore phase information were obtained in PHASER (1994) by molecular replacement using residues 0 to 143 from chain A of the AafB crystal structure (Chapter 5) as the search model in both $P4_12_12$ and $P4_32_12$. Molecular replacement was performed between resolution ranges 28.74 - 2.00 Å and only two molecules were found in the asymmetric unit in space group $P4_32_12$ with translational Z-scores of 13.5 and 10.6 and the log likelihood gain of 135 and 341, respectively (Table 7.3). Rigid-body refinement was performed in PHASER resulting in an R-factor of 0.55. Noticeably, the translational Z-score for the second molecule was lower and two clashes were found, but the electron density showed good connectivity for this region so the clashes might arise from the long loops in the AafB structure.

	Translational Z-Score	Packing Check (PAK)	Log Likelihood Gain
First Molecule	13.5	0	135
Second Molecule	10.6	2	341

Table 7.3: Z-score of the two molecules found from PHASER. Only two molecules were found in the asymmetric unit with good Z-score and log likelihood gain of 97 and 332 respectively.

The asymmetric unit of HdaB consists of two molecules (Fig. 7.1) and the crystal packing forms a well-connected lattice (Fig 7.2) with good connected density for both

protein molecules. In addition, the maps demonstrated the expected rounded density for proline residues and extended density for large side chains such as histidine, arginine, glutamic acid, glutamine and asparagine (Fig. 7.3).

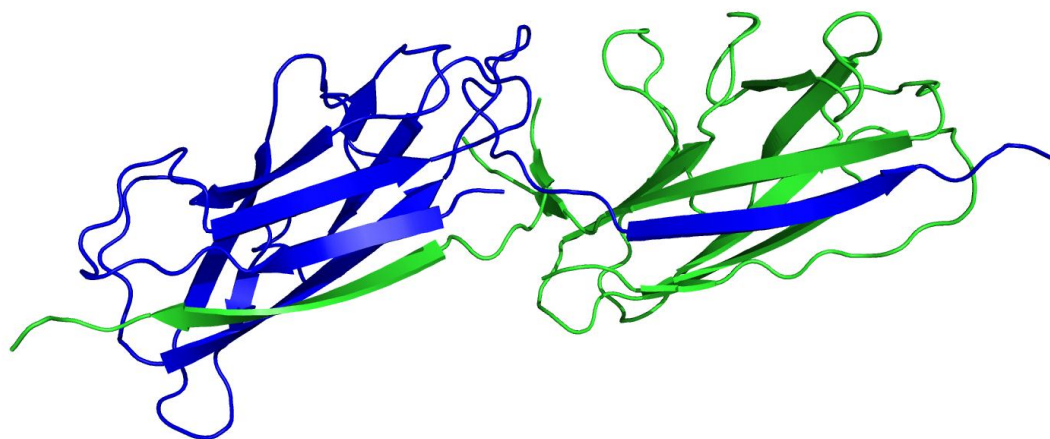


Figure 7.1: The asymmetric unit of the HdaB crystal. The HdaB asymmetric unit is made up of two molecules which are coloured green and blue.

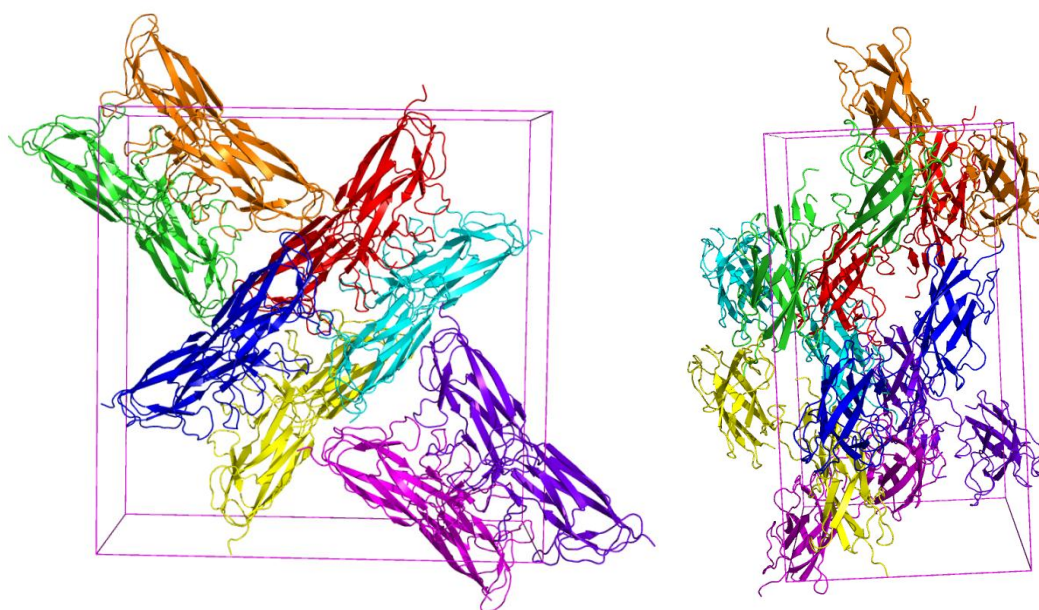


Figure 7.2: Two different views of the crystal packing arrangement within the HdaB crystal. There are 8 asymmetric units of HdaB in the $P4_32_12$ unit cell and each are coloured in a different colour.

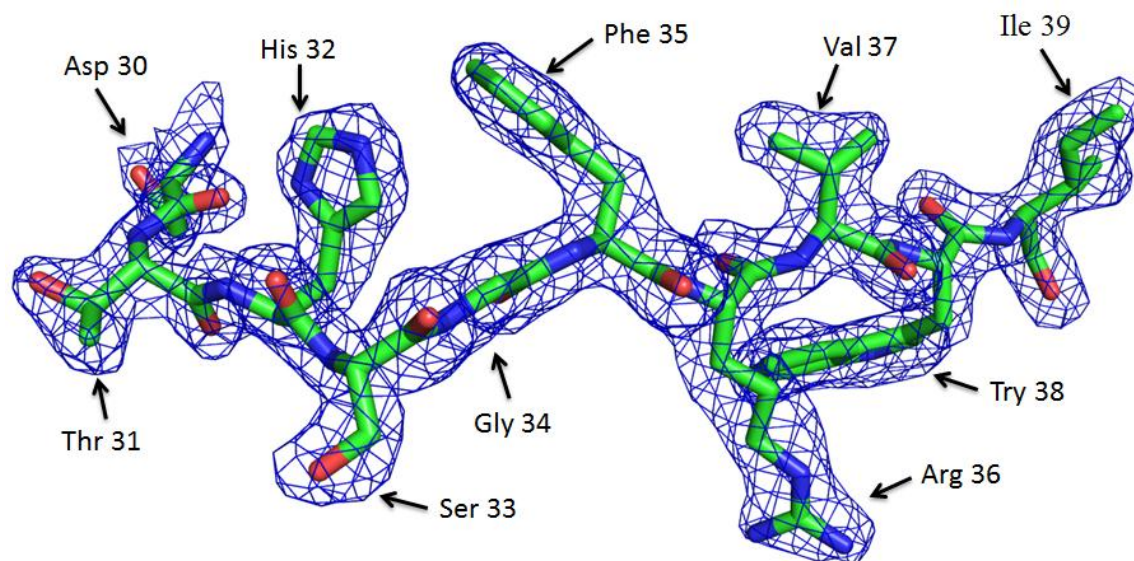


Figure 7.3: An example of the electron density map quality of HdaB. The region shown is from residue 30 to 39 of chain A. The backbone is connected with good electron density. The side chains such as arginine, histidine, phenylalanine, tryptophan, threonine, valine, serine and isoleucine were easily recognizable as shown in the $2F_o - F_c$ electron density map contoured at 1.0rms electron density. Image was generated using PyMOL.

7.4 Structural Refinement and Modelling

The engineered donor strand was based on the sequence of HdaA and a cysteine was included close to the first predicted β -strand to further stabilise the protein. As a result, a strand swapped dimer with a disulphide bond in the interface of the two domains was formed. The analysis of the HdaA sequence indicated 6 residues from the signal peptide of HdaA were included which extended the length of the original designed linker (Emanuelsson *et al.* 2007). This facilitates the flexibility of the linker and allowed the swapping of the donor strand between domains. In addition, the disulphide bond stabilises the dimer.

Positional and individual isotropic B-factor refinement were performed with the CCP4 programme REFMAC (1994). Coot was then used to mutate residues to the correct side chains, and make adjustments to the model. This was followed by cycles of refinement in REFMAC using automated 2-fold non-crystallographic symmetry

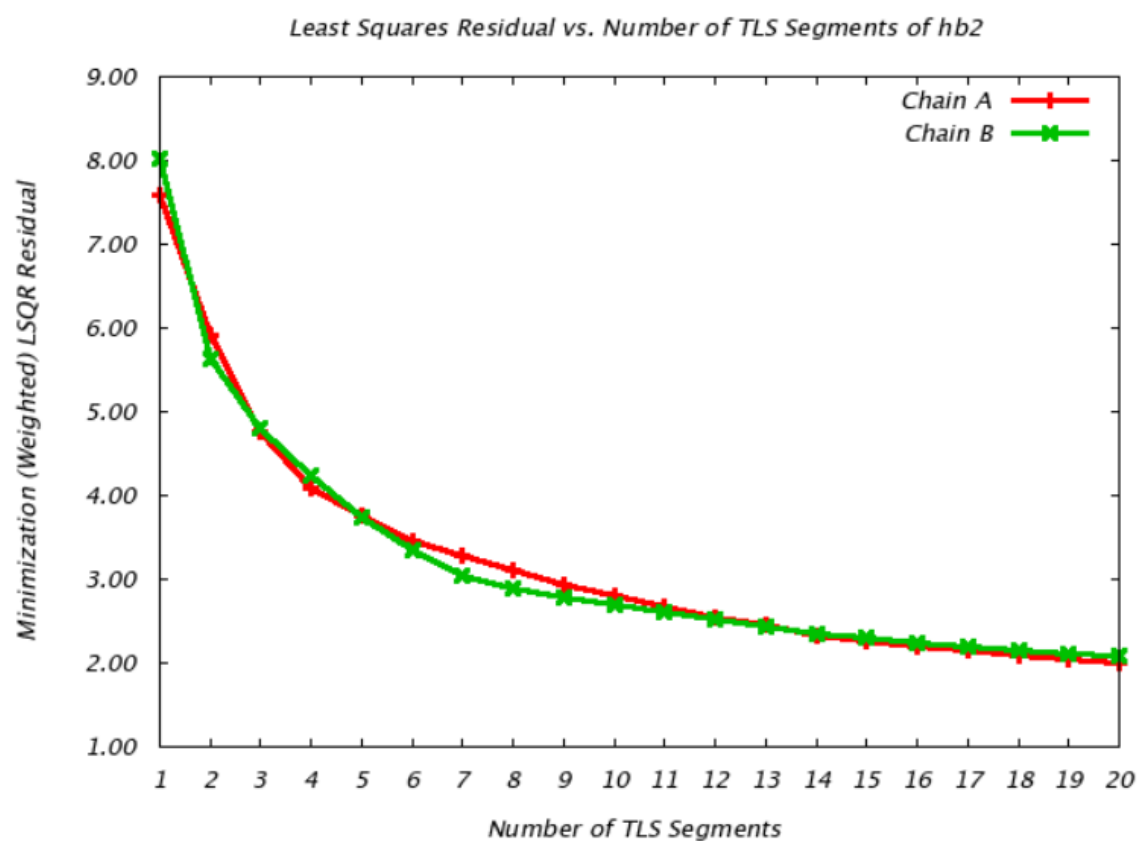
(NCS) restraints followed by model building in COOT. After many rounds of adjusting the model, the model was sent to the TLSMD (Translation/Liberation/Screw Motion Determination) server for TLS B factor analysis (Painter and Merritt 2006).

TLSMD analysis involves the definition of three parameters to describe the translational, librational and screw motions of specific segments within a structure. This provides further restraints to the model and improves the quality of the loop regions. Two segments were selected from the TLSMD server which separated the protein sequence around the point where the donor strand began and with this restraint the R-factor was reduced further (Fig 7.4). Different alternative conformations were observed between the two molecules in the asymmetric unit and different backbone positions of the loop regions were also observed therefore NCS was taken off for the last few rounds of refinement.

With the resolution of this dataset (2.0 Å), 10% of the data were removed for cross-validation during refinement. The model was refined using the 90% working data set with the aim of minimizing the value of R_{work} which was obtained after each round of refinement. To evaluate the accuracy of the model, and protect against model bias, the smaller dataset was used to calculate the R_{free} and was expected to correlate with R_{work} . As in this case, the model had similar R_{work} and R_{free} which suggested no bias in the model and with no possible adjustment to the protein model, based on inspection of the electron density maps and no further reduction in the R-factors, a search of ions and water molecules was performed. This resulted in 318 water molecules, three acetate ions and one iodide ion. The statistics of the final refinement are given in Table 7.4.

Refinement Statistics		Values
Refinement Resolution (Å)		28.74 - 2.00
Reflection used in Refinement		23905
Reflections used in R_{free} Calculation		2654
Total Protein Residues in Asymmetric Unit Model		527
Number of Protein Molecules in the Asymmetric Unit		2
Number of Ions		4 (3 citrate ion and one iodide ion)
Number of Water Molecules		318
Number of Non-hydrogen Atoms		2613
RMS Deviation from Ideal	Bond Length (Å)	0.009
	Bond Angles (°)	1.184
	Dihedral Angles (°)	6.142
R_{work}		0.20733
R_{free}		0.24831

Table 7.4: Refinement statistics for the final refined HdaB model.



Chains Alignment using 2 TLS Groups

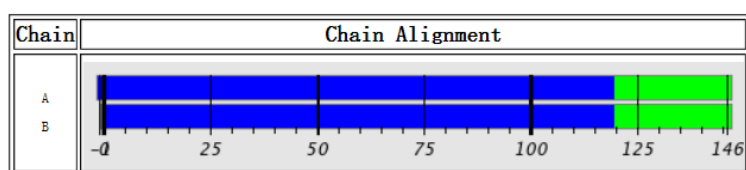
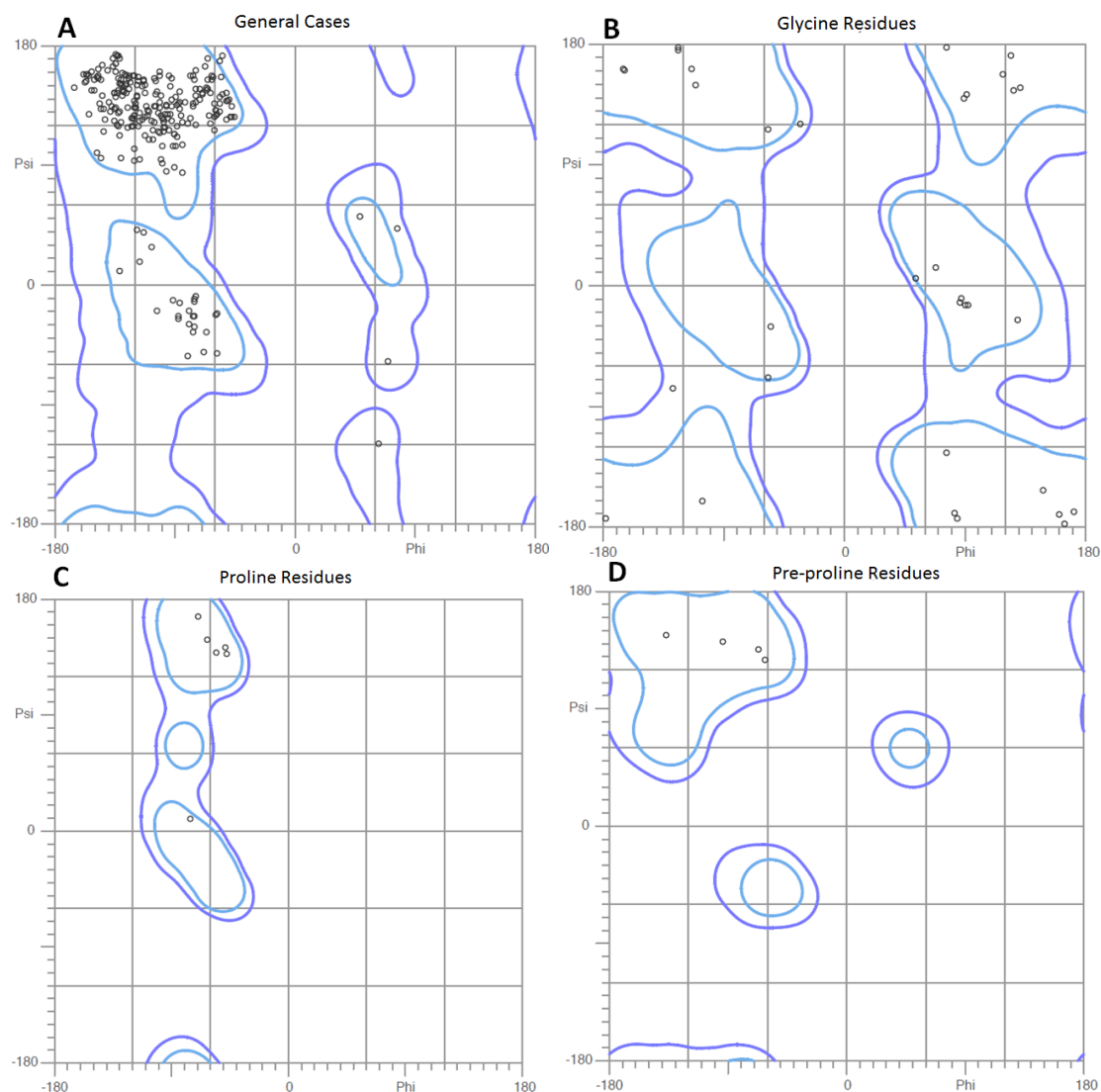


Figure 7.4: Results from TLSMD. Two segments were used for TLSMD restraints, in which it divided the protein into two, the native protein and the donor stand.

7.5 Model Analysis and Validation

With no further refinement or model building leading to improvements to the model, the structure was validated using COOT, MOLPROBITY (Chen *et al.* 2010) and the PDB server. All atoms in the backbone and side chains were surrounded by good electron density and orientation of each side chain was adjusted by rotamer analysis. With the resolution of the data, both the bond lengths and bond angles were weighted less than theoretical values. Furthermore, no free electron density blobs were discovered at this stage which suggested no additional ligand or water molecules were present. The Ramachandran analysis reported by MOLPROBITY gave 97.7% (291/298) of the residues resident in the favoured regions and just 2.3% (7/298) of the residues falling within the allowed regions (Fig. 7.5).



Residues in most favourable regions	291	97.7%
Residues in allowed regions	7	2.3%
Residues in disallowed regions	0	0%
Number of non-glycine, non-proline and pre-proline residues	254	100%
Number of glycine residues	34	
Number of proline residues	4	
Number of pre-proline residues	6	
Total number of residues	298	

Figure 7.5: Ramachandran plots for the structure of HdaB. The areas inside the light blue lines represent the favoured regions and the areas within the purple blue regions represent the allowed regions. Image generated from MOLPROBITY (Chen *et al.* 2010).

The final model contains two protein chains (25 residues have alternative conformations), three citrate ions, one iodide ion and 318 water molecules. The N-terminal His-tag of HdaB was not visible in the electron density maps. Chain A begins at residue -1 to residue 146 (although residue 124 was not modelled), whilst chain B begins at residue 0 to residue 146. The missing residue was part of a loop and was adjacent to residues with high B factors (Fig. 7.6). The overall Wilson B-factor for HdaB is $29.53 \text{ \AA}^2$ which correlates well with the final $\langle B \rangle$ factors for the final structure. Interestingly, β strand number 8 and 9 are only present in chain A and are located on the engineered linker region but both chains display high B factors. This suggested these two β strands are relatively flexible.

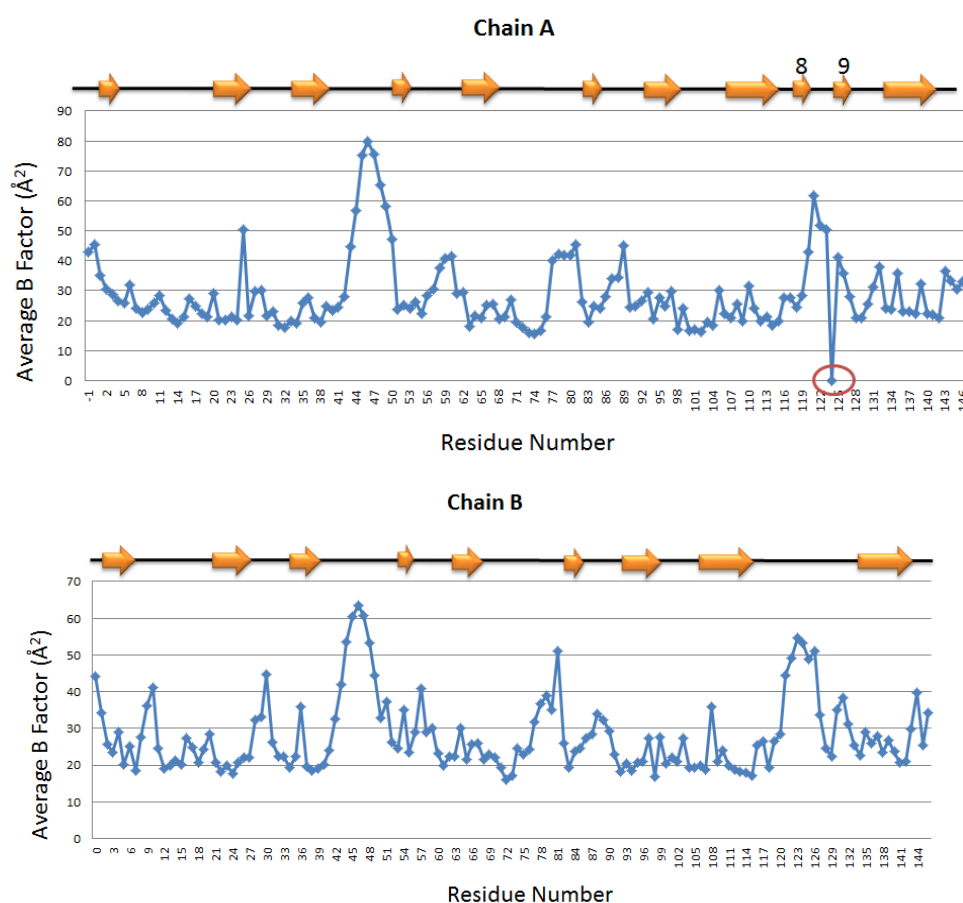


Figure 7.6: Average B factor of HdaB residues. The average B factors were plotted for both chains with the value of average B factor on the Y-axis and the residue number on the X-axis. The red circled area indicates the missing residues and the black lines and the orange arrows above represent the secondary structure for the loops and β -strands, respectively. Both β strand 8 and 9 in chain A are numbered.

7.6 Structural Overview of HdaB

7.6.1 Structure of HdaB

The immunoglobulin-like fold of this protein is similar to other bacterial pilus. The secondary structure of HdaB displays two layers of β sheets which consist of 9 (chain B) and 11 (chain A) β strands including the engineered donor strand on the C-terminus (Fig. 7.7). The two additional short (3 residues) β strands on chain A are part of the engineered linker which has higher B factor values (above the average B factor) and is very near to the missing residue Lys 124, therefore the presence of this β strand structure is probably dependent on crystal contacts. Chain B will be used for further comparison with homologous structures.

An additional cysteine was engineered to the linker. That formed disulphide bond which strengthened the connection in the asymmetric unit and may have improved the crystallisation. Initially, the design of the donor strand was to promote self-complementation to produce a stable monomer but by adding this additional cysteine, the presence of the disulphide bond has facilitated the exchange of the donor strand, where the strand is flipped out and inserted into the groove of the adjacent molecule as similar to the chaperone-subunit and subunit-subunit complexes.

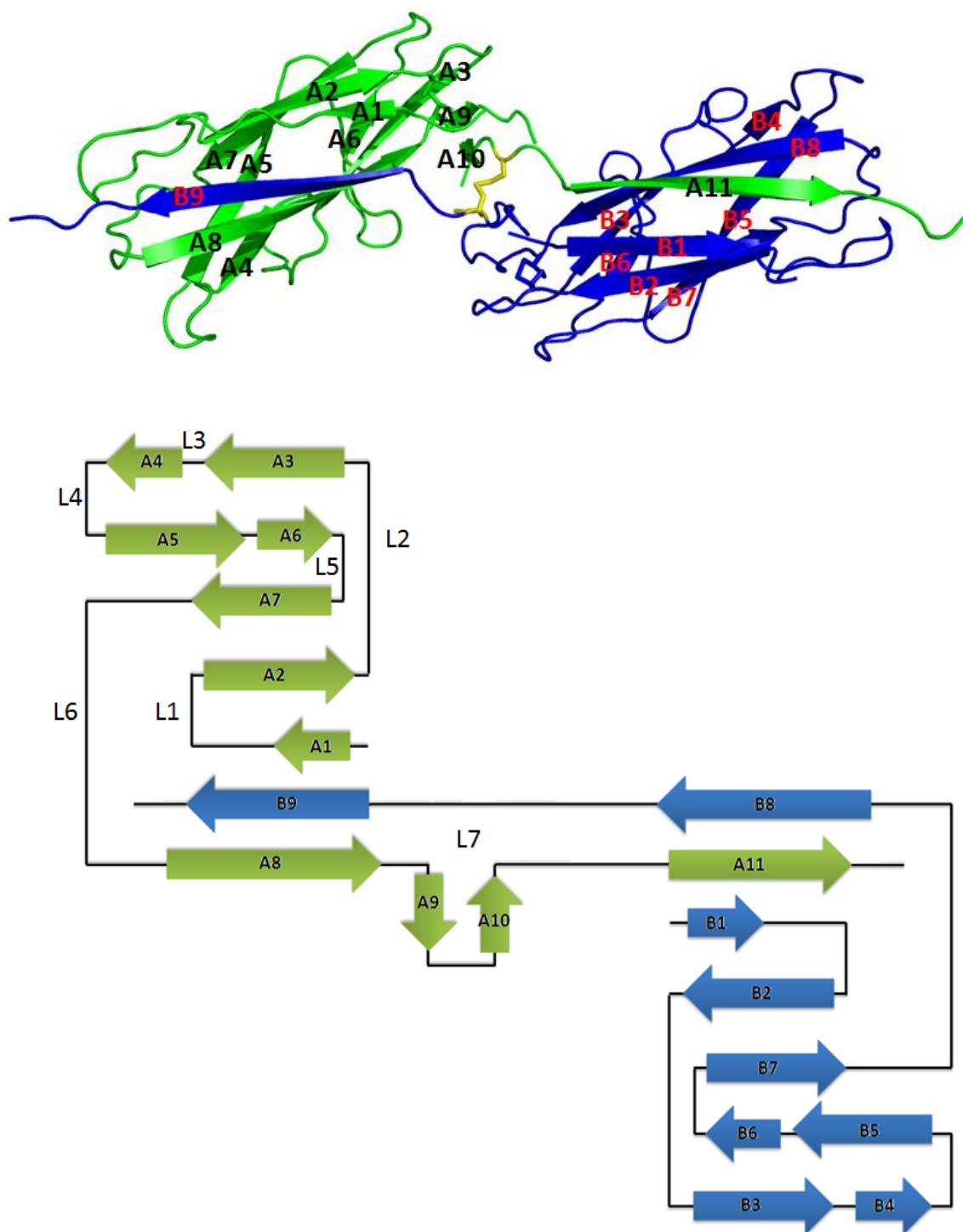


Figure 7.7: Ribbon representation and topology of the Structure of HdaB. The additional cysteines are coloured in yellow with stick representation. One molecule is coloured in green and the other is in blue. Each strand is numbered in the order of sequence, and some loops are also labelled for further discussion in this chapter.

7.6.2 Comparisons of Molecules within the Asymmetric Unit

There are two molecules present in the asymmetric unit and there was used to provide further restraints via NCS to improve the quality of the electron density map during processing. Despite the backbone of these molecules being very similar (rmsd of 0.695 Å), there were some differences in the loop regions and in the length of some β -strands (Fig. 7.8). The loops labelled in Figure 7.7 (L1-L7) displayed large differences in position especially loops 3, 4 and 7 (L3, L4 and L7). The three areas with most disagreement between the two molecules are located around residue Lys 124 and are flanked by residues with high B factors. This suggests that there is high flexibility in these regions due to the lack of contact between adjacent molecules (Fig 7.9). Furthermore, a region of loop 7 (L7) in chain B is different to the corresponding region in chain A where two short β strands are formed as a result of the engineered donor strand. In addition, the engineered donor strand might be too long, which has allowed the donor strand exchange mechanism to form a dimer. The inter-molecular disulphide bond may also cause the difference in the position of the donor strand between the two molecules. This might slightly alter the position of the protein tertiary structure hence affecting the crystal lattice formation and contacts between molecules which leads to different position of the loops.

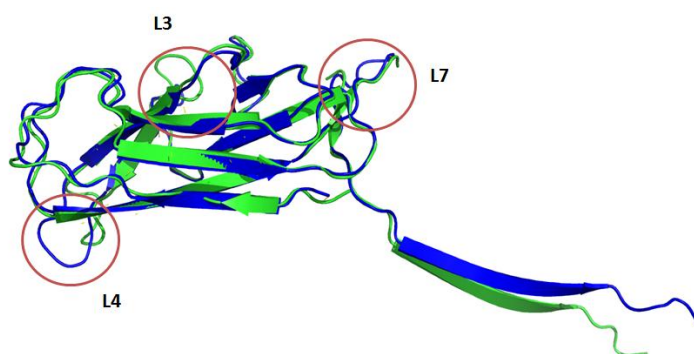


Figure 7.8: Superimposition of the two molecules in the asymmetric unit. One molecule is in green and the other is in blue. The overall structure was very similar with all the β -strands overlaid on each other, but differences are circled and labelled L3, L4 and L7.

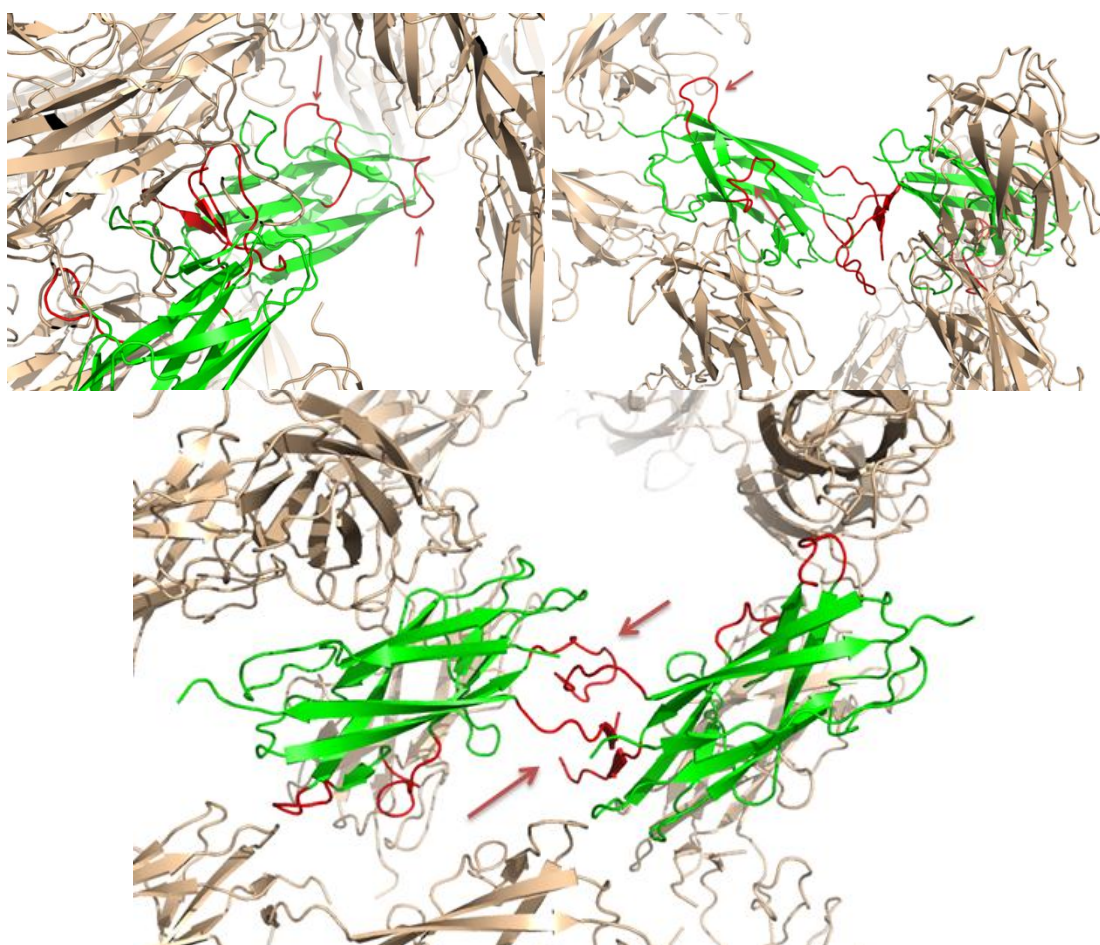


Figure 7.9: Adjacent contacts around the asymmetric unit. The centre asymmetric unit is coloured in green with the regions of interest in red (L3, L4 and L7); and the adjacent asymmetric units are all in brown. The red arrows indicate the loop regions of disagreement where there are no contacts to any adjacent molecules.

Although the crystals contain a rather high solvent content (56.99%) and large solvent channels running through the crystals, each asymmetric unit shows good contacts to adjacent asymmetric units (Fig . 7.10).

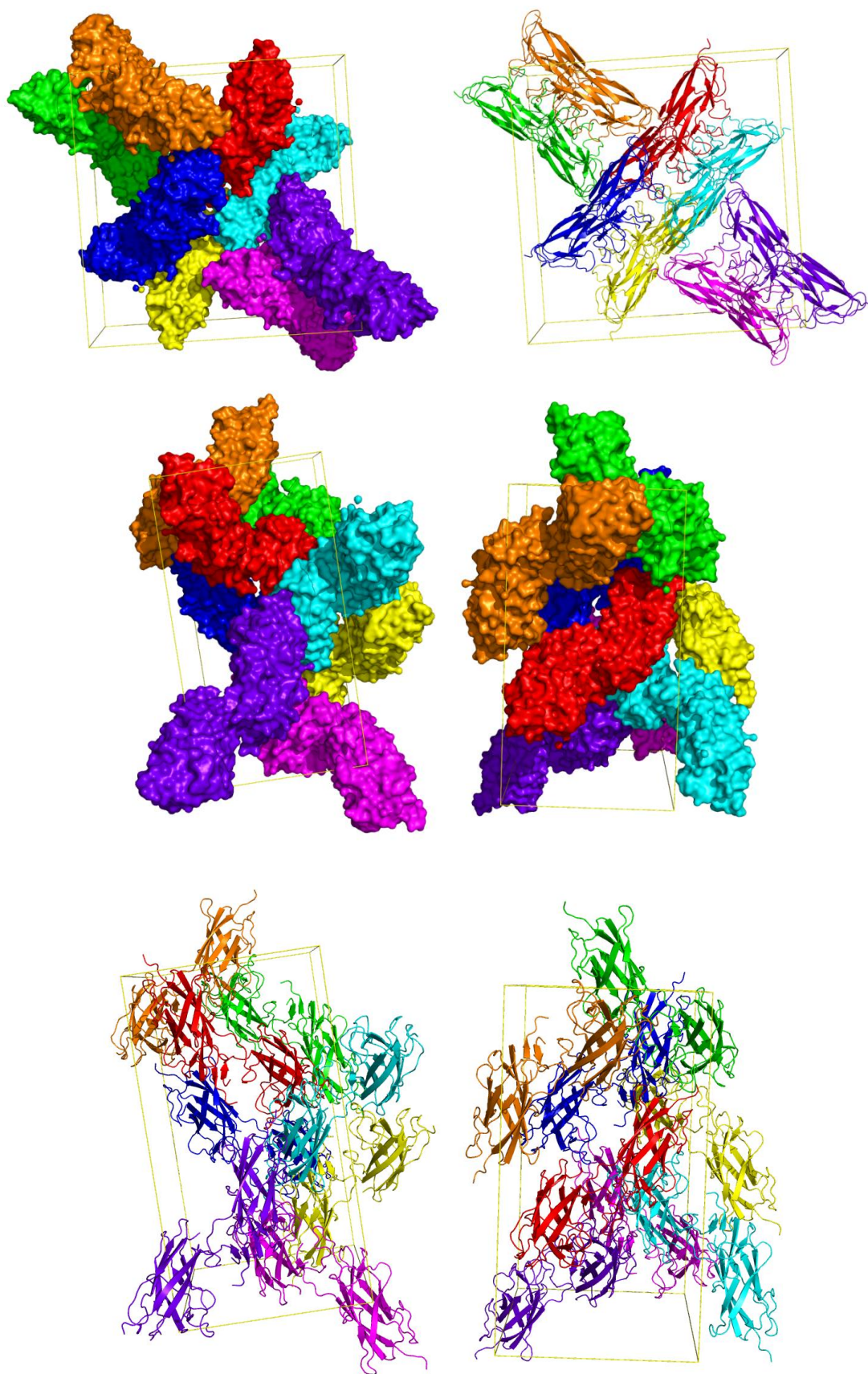


Figure 7.10: Crystals contacts between asymmetric units in the HdaB unit cell. Different views of the contact between asymmetric units with different colours representing each asymmetric unit.

7.6.3 Potential Binding Sites

There have been no previous studies on the binding of Hda fimbriae but other homologous proteins such as AafB exhibit the potential to bind negatively charged molecules (chapter 6). The β -strands with the three residues involved in the putative interaction between AafB and heparin are compared to HdaB. The sequence alignment indicates many conserved and similar residues (charged or hydrophobic side chains) presence around these regions which suggest the present of similar binding sites (Fig. 7.11A). The structure comparison also shows the side chains of Thr 24 (AafB) (Thr 23 in HdaB) and His 37 (AafB) (Arg 36 in HdaB) have an orientation that might facilitate ligand binding (Fig. 7.11B). However, these residues could also be affected by conformational changes on binding rather than involving in direct contact in binding.

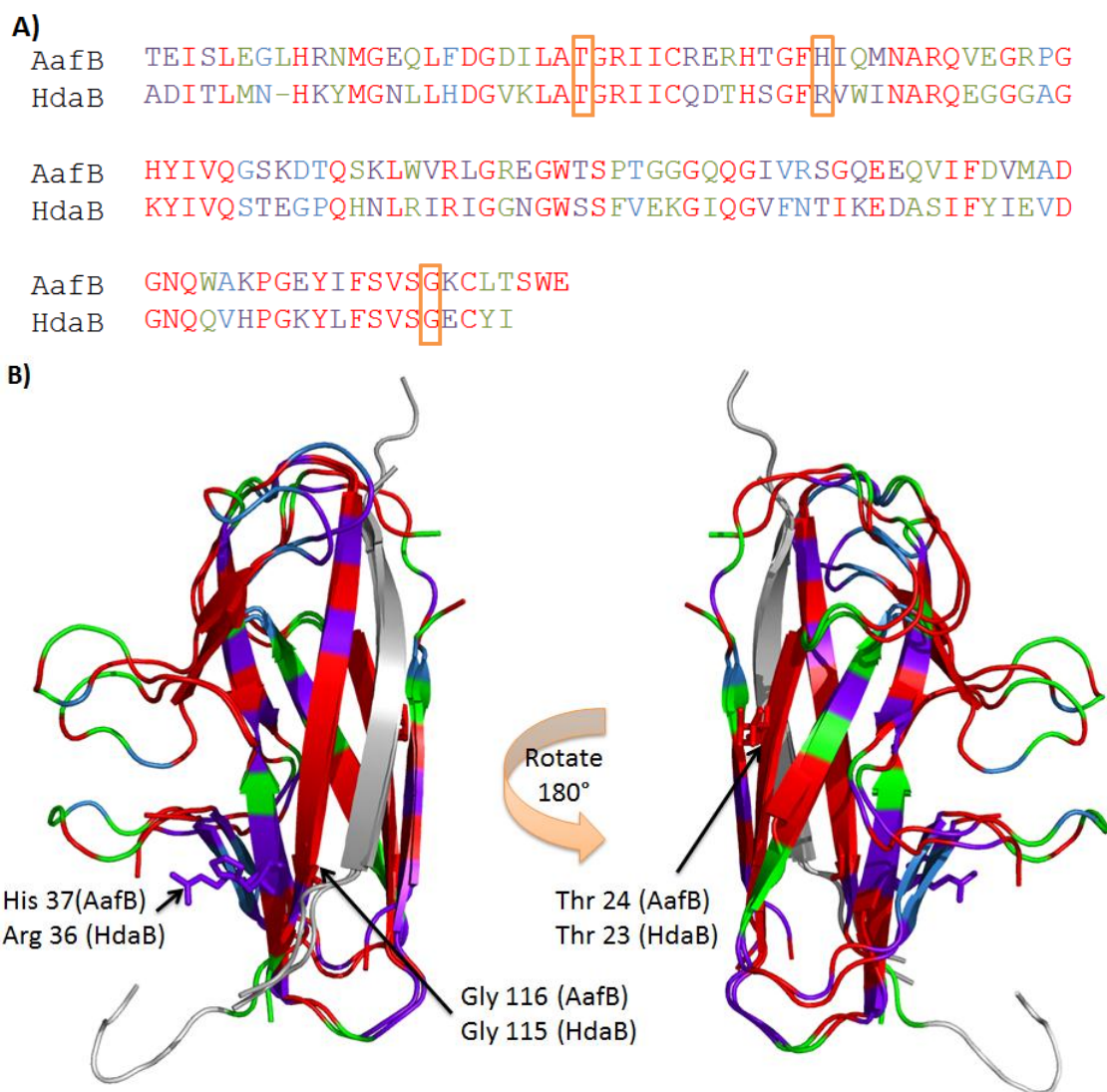


Figure 7.11: Alignment of AafB with HdaB. The three residues identified in the putative interaction between AafB and heparin are indicated by orange box/black arrow. The same residues on the strands are coloured red and the similar residues are in purple. The less similar residues are in blue and different residues are in green. A) Sequence alignment of AafB with HdaB. B) Superimposition of AafB with HdaB. The donor strands are not compared and are coloured in grey.

The HdaB structure has many regions of negatively and positively charged surface and from the ligands which have been modelled in the structure, suggesting potential binding sites (Fig. 7.12). In particular, three citrate ions were found in contact with HdaB and water molecules. Whilst these may be artefacts of the crystallisation process, they also indicate putative binding sites and need further investigation (Kumar *et al.* 2010).

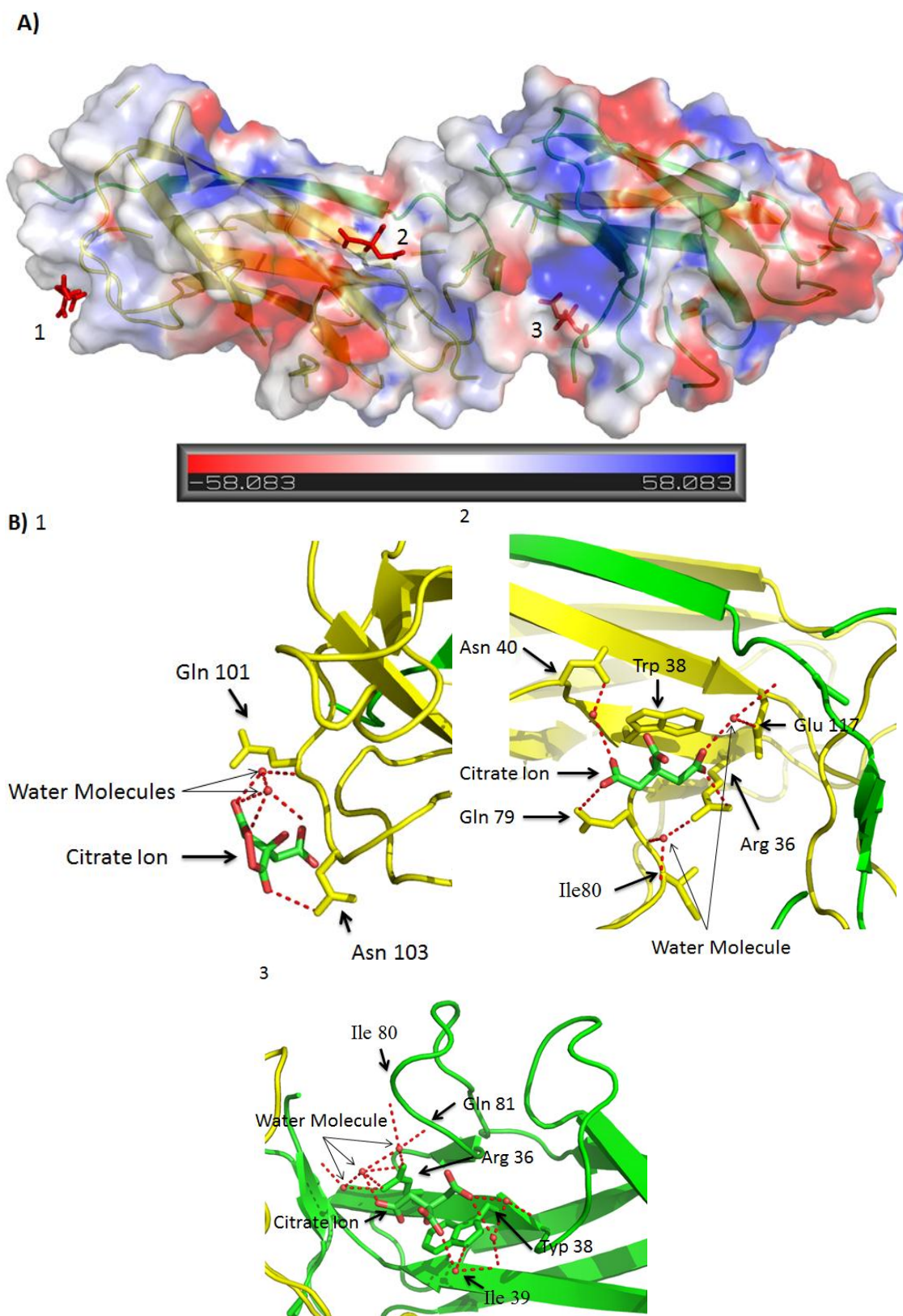


Figure 7.12: The contact regions between citrate ions and HdaB. The structure of HdaB is coloured in yellow (chain A) and green (chain B). A) The electrostatic surface of HdaB. The areas shaded red and blue are negatively charged and positively charged, respectively. Beneath the surface is the ribbon representation of the structure of HdaB. The citrate ions are showing red sticks. B) The areas involved in the contact with citrate ions (1, 2 and 3). The hydrogen contacts between citrate ion, water molecules and the corresponding residues are linked by red dash lines. The citrate ions are represented in sticks with the colour green and the oxygen atoms at the COO⁻ groups are in red.

7.7 Comparisons of HdaB with Afa/Dr/Aaf Homologues

7.7.1 Sequence Alignment

Two known structures from the Afa/Dr family and AafB were used for the comparisons with HdaB (Fig. 7.13). The additional engineered donor strand sequences were removed in this alignment hence only the native protein sequences were used. The sequence alignment shows that AfaD and DraD are identical and many residues are conserved between AafB and HdaB. However, the phylogram (Fig. 1.16) suggests that AafB is more closely related to Afa/Dr than HdaB despite them being from the same strain of *E. coli* (EAEC). The similarities and differences were explored further by secondary and tertiary structure comparison.

```
AfaD      AELHLE-SRGGSGTQLRDGAKVATGRIICREAHHTGFHVWMNERQVDGRAERYVVQ-SKDG 58
DraD      AELHLE-SRGGSGTQLRDGAKVATGRIICREAHHTGFHVWMNERQVDGRAERYVVQ-SKDG 58
AafB      TEISLEGLHNMGEQLFDGDILATGRIICREAHHTGFHIQMNARQVEGPGPHYIVQGSKDT 60
HdaB      ADITLM-NHKYMGNLLHDGVKLATGRIICQDTHSGFRVWINARQEGGGAGKYIVQSTEGP 59
          :: * : * * * :*****:: *::: : * * * . :*:* ::.

AfaD      RHELVRVTGGDGWSPVKGEGGKGVSPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVV 118
DraD      RHELVRVTGGDGWSPVKGEGGKGVSPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVV 118
AafB      QSKLWVRLGREGWTSPTGGGQQGIVRSGQEEQVIFDVMADGNQWAKPGYIFSVSGKCLT 120
HdaB      QHNLRIIRIGNGWSSFVEKGIQGVFNTIKEDASIFYIEVDGNQQVHPGKYLFSVSGECYI 119
          : : * : * * :*:* . * : * : . : * : : * * * * * * * * *

AfaD      PQE 121
DraD      PQE 121
AafB      SWE 123
HdaB      ---
```

Figure 7.13: Sequence Alignment of HdaB with Afa/Dr/Aaf. The amino acids are coloured differently for easy observation of any disagreement between the sequences and the star under the residues indicated the present of identical amino acid at the position.

7.7.2 Secondary Structure Comparisons

AafB (see Chapter 5) and HdaB both have 9 β -strands and both AfaD and DraD appear to have more β -strands in their secondary structure. However, by comparing the regions of the β -strands in these four proteins, the positions of the conserved residues are very similar. The additional β -strands in AfaD and DraD were either part of a long β -strand in AafB and HdaB or located at the loop regions (Fig. 7.14). This suggests that these four proteins have the same number of β -strands.

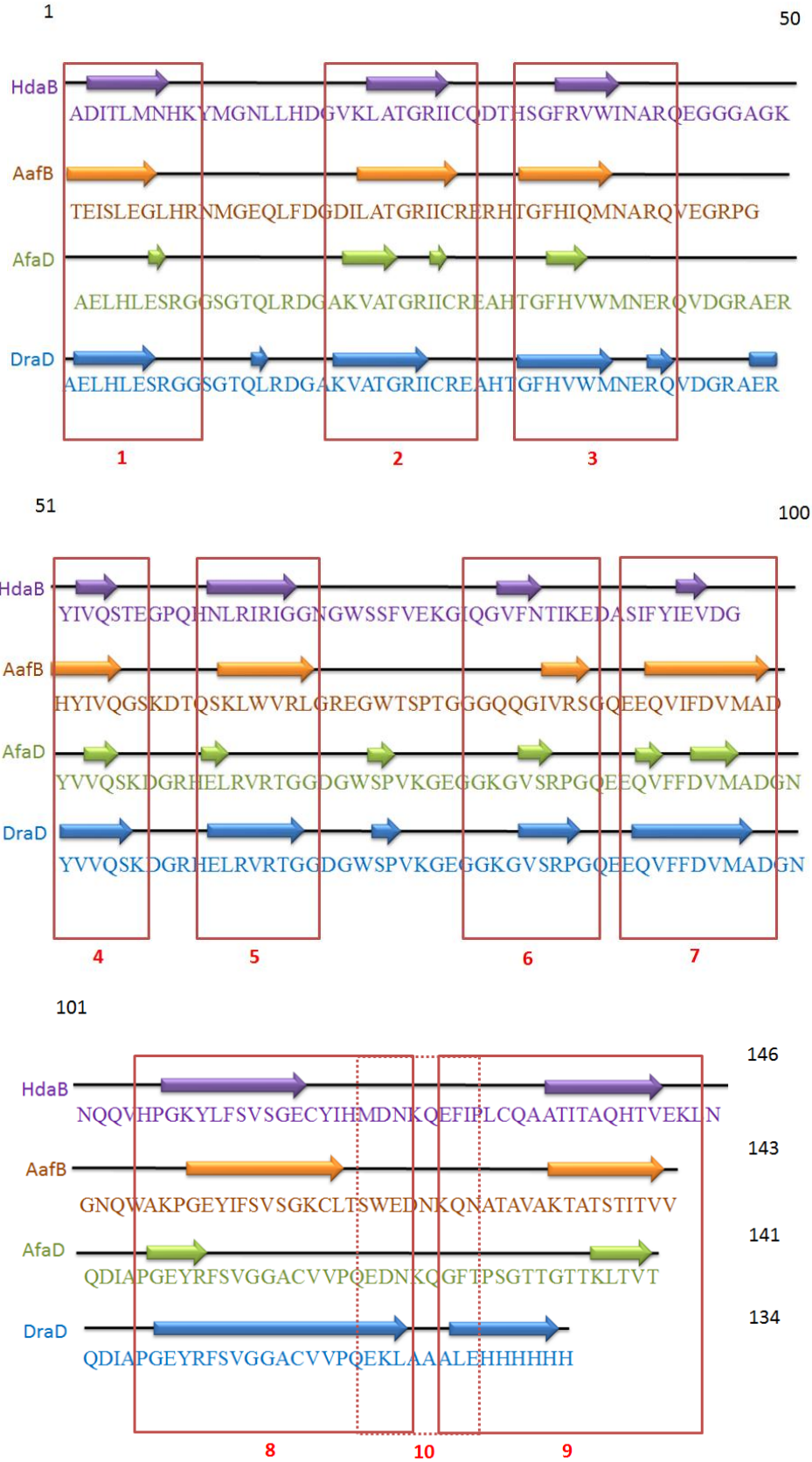


Figure 7.14: Comparisons of the secondary structure of HdaB with Afa/Dr/AafB. The arrows represent β -strands and the red areas with numbers are the approximate locations of the conserved β -strands. AafB is in orange, AfaD is in green, HdaB is in purple and DraD is in blue. The dotted areas were the disagreement region between chain A and Chain B of HdaB (area 10).

7.7.3 Tertiary Structure Comparisons

The structure of HdaB was superimposed on those of AfaD (PDB: 2FVN), DraD (PDB: 2AXW) and AafB. The position of backbone atoms is very similar with rmsds of 2.424 Å, 0.879 Å and 1.042 Å, respectively (Fig. 7.15). However, differences are observed in most loop regions.

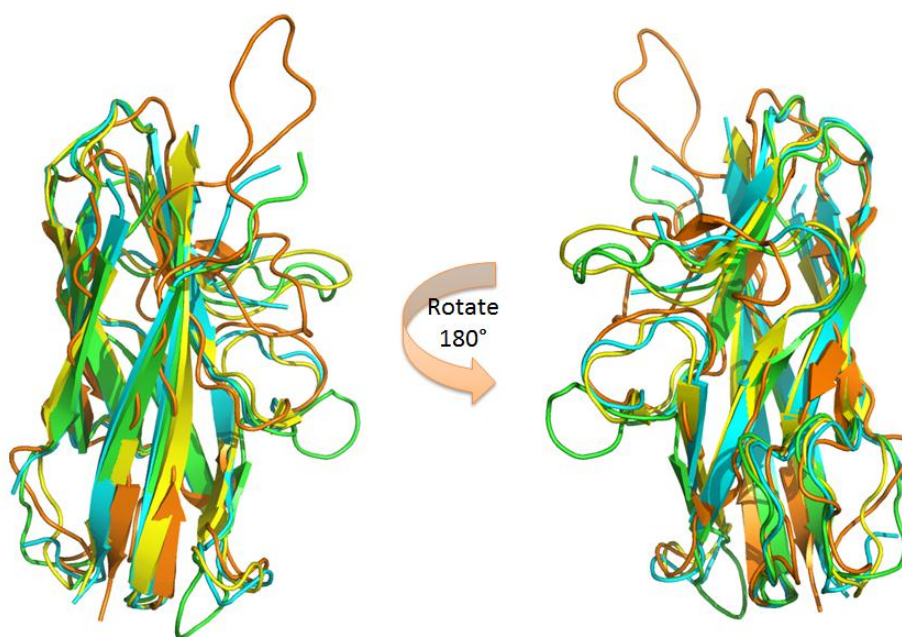


Figure 7.15: Superimposition of HdaB with AfaD, DraD and AafB. AfaB is coloured in cyan, AfaD is in orange, DraD is in yellow and HdaB is in green. Overall the structures were similar with some differences on the loop regions.

Separate comparisons between HdaB with AafB, DraD and AfaD were also carried out to closely examine the differences. HdaB is very similar to AafB and DraD as almost all the β -strands aligned perfectly and some loops were either at the identical position or followed the same orientation and directions. Five regions with the most disagreement between the structures occurred on the loops (Fig. 7.16). Region 1, 4 and 5 are the regions where the missing residues in AafB structure indicate a high degree of flexibility of these loops. In addition, region 5 is the location of the engineered linker for the donor strand and both HdaB and AafB have a different

linker to DraD which explains the differences. Although HdaB and AafB share a similar linker, AafB was able to self-complement and form a monomeric species, whilst HdaB forms a dimer; hence the differences in the linker are to be expected. Furthermore, despite all the residues being modelled in region 2 and 3, the high B factors in both AafB and HdaB suggest these residues have a greater freedom of movement.

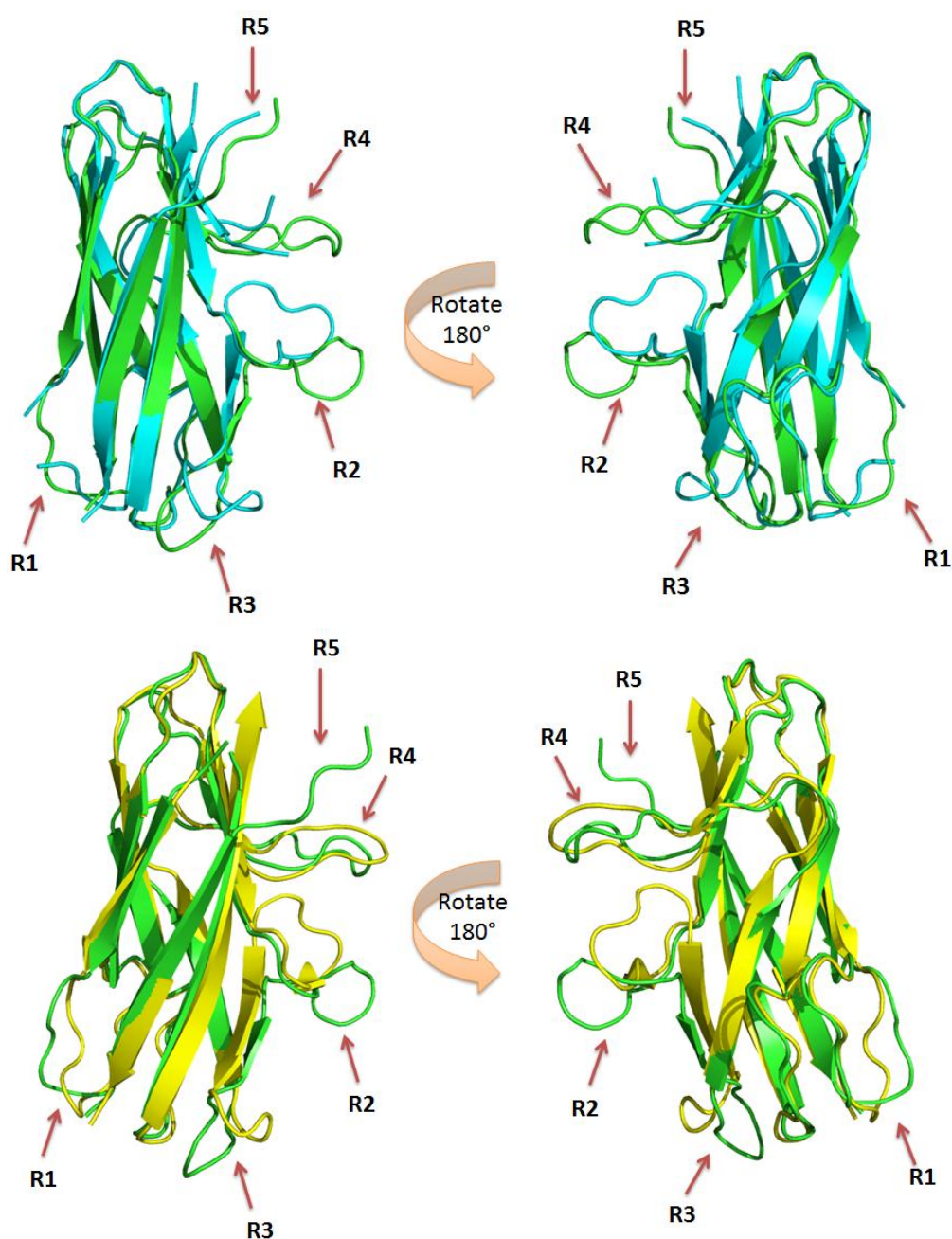


Figure 7.16: Superimposition of HdaB with DraD and AafB. AfaB is coloured in cyan, DraD is in yellow and HdaB is in blue. The red arrows indicate the regions with most differences between the structures.

Likewise, HdaB and AfaD have a similar overall fold, but the positions of many loops and some β -strands diverge. Many short β -strands in AfaD are part of the longer β -strands in HdaB with some minor differences (Fig. 7.17). The loops are particularly different between the two structures which can be explained by high B factors in these

residues of HdaB and the AfaD structure was determined by solution NMR in which the sample would be more flexible.

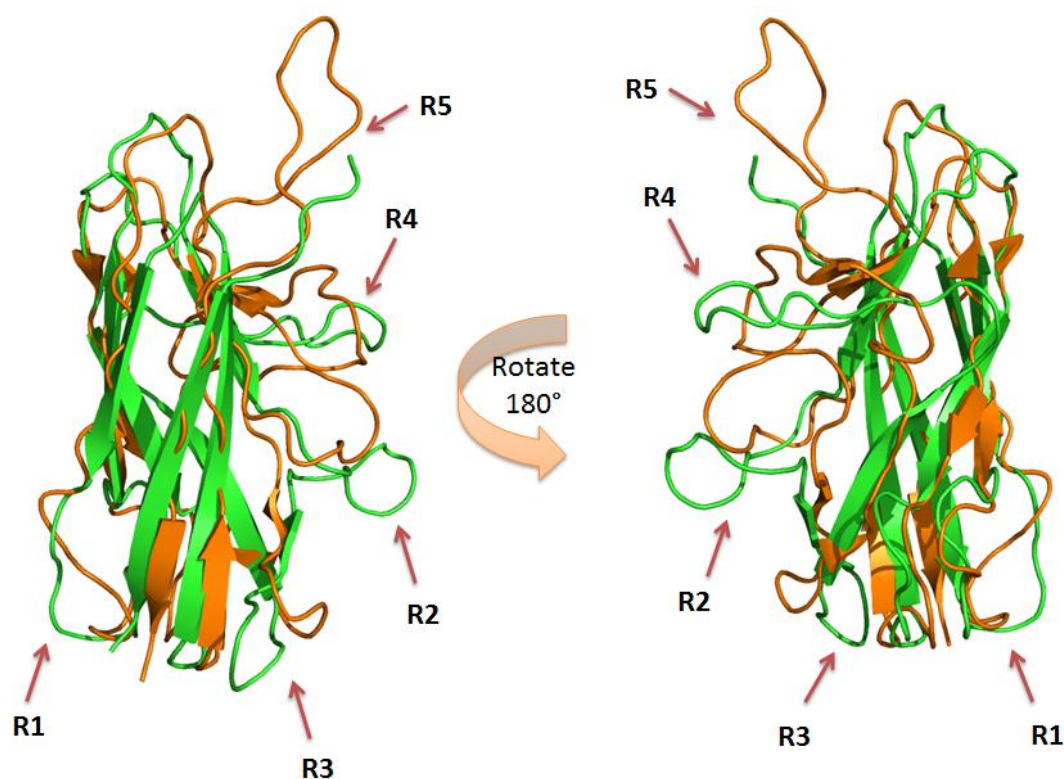


Figure 7.17: Superimposition of HdaB with AfaD. AfaD is coloured in orange and HdaB is in green. The red arrows indicate the regions with most different regions in the structures.

7.8 Summary

The structure of HdaB was solved by X-ray crystallography to a resolution of 2.0 Å. Chain A consists of 11 β -strands and chain B consists of 9 β -strands. The space group is $P4_32_12$ and the final $R_{\text{work}}/R_{\text{free}}$ is 0.20733/0.24831. All the residues fall within the favoured or allowed regions of the Ramachandran plot. Two molecules were found in the asymmetric unit which are very similar but with some differences around the loops (especially L3, L4 and L7). Both L3 and L4 consist of residues with high B factors with no direct contacts to other adjacent molecules. L7 is part of the engineered donor strand and an additional cysteine was placed in the linker for additional disulphide bond formation, which might have caused differences in this

region and the position of the donor strand. However, the disulphide bond facilitated the formation of a dimer which might have aided crystallization.

Only one iodide ion was sufficiently ordered to model it, with an occupancy of 0.3 which explains the lack of anomalous signal for phasing. Three citrate ions were found interacting with HdaB which facilitate the overall crystal packing by establishing more contacts between symmetric units. These regions of the protein are positively charged and could potentially function as binding sites. There has been no binding site reported to date for HdaB but microarray screening and NMR studies in this project have suggested potential binding with oligosaccharides such as chondroitins and heparin for AafB. This would be a logical direction for investigation of the function of HdaB.

Comparisons with the sequence of AfaD, DraD and AafB predicted a similar immunoglobulin-like fold for HdaB. Both DraD and AafB are structurally similar to HdaB with differences on several loops which can be explained by the high flexibility of these regions. In addition, AfaD contains more β -strands than HdaB and AafB, but as the comparison has shown, most of these extra short β -strands correspond to the longer β -strands in HdaB and AafB. Furthermore, HdaB and AfaD do show minor differences in the positions of several β -strands.

Chapter 8: Structural Studies of SefD

8.1 Data Collection

A single large crystal of SefD was quickly passed through cryoprotectant (composition of cryoprotectant is given in section 3.4.3) and then flash frozen in liquid nitrogen. This was taken to beamline I24 (Diamond Light Source, UK) and exposed to synchrotron radiation with wavelength 0.97780 Å where it diffracted to 3.0 Å. After indexing 2 images separated by 90°, the space group was inconclusive. Therefore 360° of data were collected at 100 °K with oscillation angle of 0.2° per frame to 3.0 Å on a Pilatus 6M detector, with a detector distance of 298 mm.

8.2 Data Processing

The data were processed and indexed in the point groups P622 and C222 to 3.0 Å using MOSFLM (1994) and scaled with SCALA (1994). The results from SCALA were fed into Pointless (1994) which again suggested the possibility of point group P622 and C222, but was unable to determine the true space group. The statistics for processing and scaling of the SefD data are given in Table 8.1. Therefore, the data was reprocessed, scaled and truncated in all derivative space groups. Both point groups were analysed for the crystal composition and with this resolution, the solvent concentration is likely to be reasonably high and hence one molecule in the asymmetric unit is expected for point group P622 and 3 molecules for C222 (Table 8.2).

Protein		SefD (with SefA DSC strand)	
Point group		P622	C222
Cell dimensions (Å)	a	52.2499	52.2585
	b	52.2499	90.47
	c	216.8550	216.8473
Cell angles (°)	α	90.00	90.00
	β	90.00	90.00
	γ	120.00	90.00
Number of reflections (total)		20815	49778
Number of reflections (unique)		4422	10689
I/ σ I (mean)		4.9 (1.3)	3.8 (1.3)
Multiplicity		4.7 (5.2)	4.7 (5.0)
R _{sym} (%)		11.8 (60.2)	15.1 (55.8)
Completeness (%)		99.7 (99.9)	99.7 (100)
Resolution range (Å)		45.235 – 3.0	45.235 – 3.0
Average B Value (Å ²)		74.246	68.196

Table 8.1: Summary of data collection and scaling statistics for SefD. The data were processed in the point group of P622 and C222. The statistics for the outliers are in bracket.

Point Group	Number of molecules per asymmetric unit	Matthews Coefficient (Vm)	Solvent Content (%)
P622	1	2.67	53.97
	2	1.34	7.93
C222	1	8.01	84.65
	2	4	69.31
	3	2.67	53.96
	4	2.0	38.61
	5	1.6	23.26
	6	1.33	7.92

Table 8.2: Summary of Matthews Coefficient for SefD crystal. The number of molecules per asymmetric unit is given with the corresponding Matthews coefficient and solvent content.

A Patterson self-rotation function was performed using MOLREP for both C222 and P622. For both point groups, the Patterson self-rotation function indicates a strong 2-fold axis parallel to the a , b and c axis and 3-fold and 6-fold symmetry along the crystallographic c -axis (Fig. 8.1). In addition, the systematic absences for C222 also suggested a 6-fold symmetry along the c axis (Fig. 8.2). These suggested a point group of P622, with a 6-fold screw axis along c is more likely.

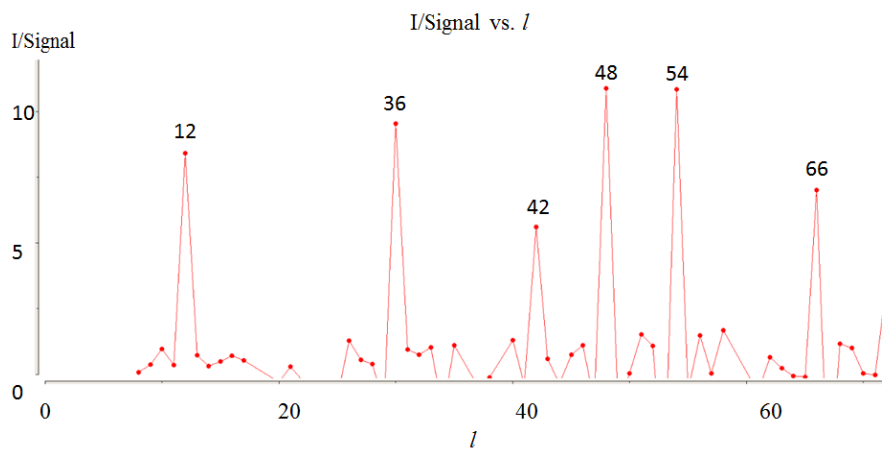
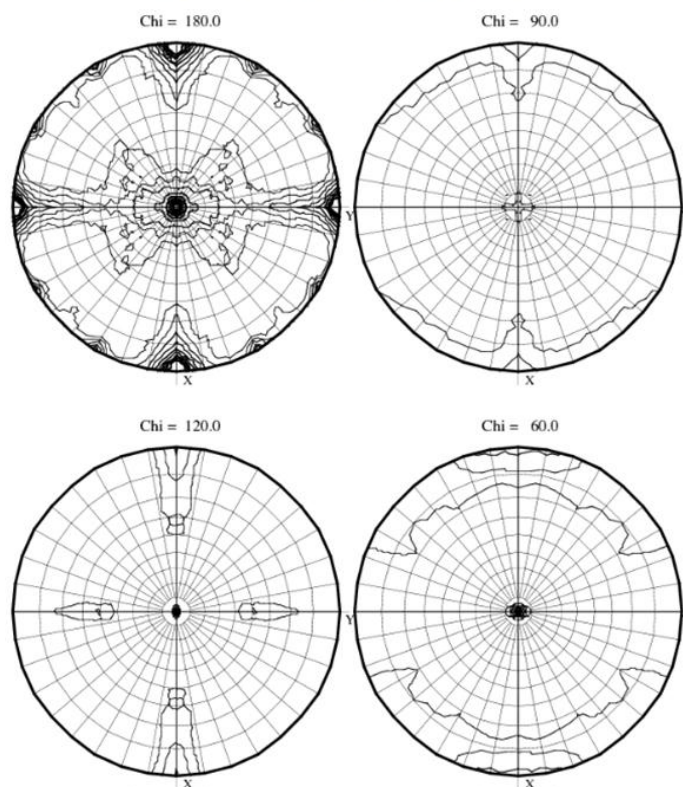


Figure 8.1: Systematic absences along the $00l$ reflections for C222. I/Signal against l demonstrates strong intensity on the positions with intervals of 6 which indicates a 6_1 or 6_5 screw axis along the crystallographic c -axis ($l=6n$), although $6_2/6_4$ ($l=2n$) symmetry is still possible.

Self Rotation Function of C222



Self Rotation Function of P622

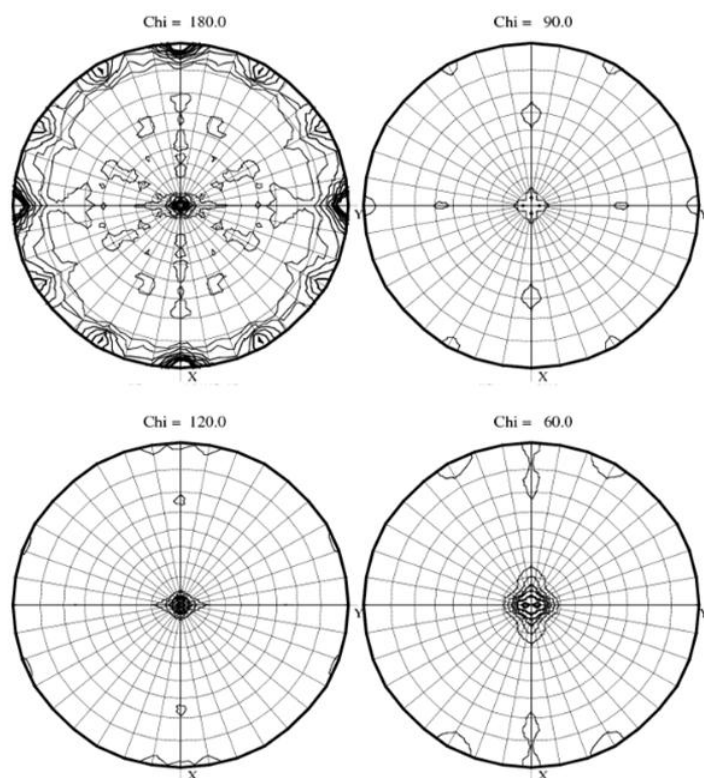


Figure 8.2: Self-rotation function of SefD crystals. This indicates strong 2, 3 and 6 fold symmetry as the strong peak in the centre of $\text{Chi}=180^\circ$ and 60° . In addition, a two 2-folds are present along the *a* and *b* axis as is observed in $\text{Chi}=180^\circ$.

Phases were calculated with PHASER (1994) using chain A of the AafB, HdaB and DraD crystal structures for molecular replacement. Molecular replacement was performed between resolution ranges 45.235 - 3.0 Å with all possible space groups but only a few partial solutions were found in three space groups with low Z-scores (Table 8.3). The model of the structure was viewed in Coot where non-continuous electron density was observed hence these partial solutions were rejected. In addition to the low sequence similarity of SefD to AafB, HdaB and DraD (below 20%), the protein sequence of SefD consists of no cysteines which might be the cause of the disagreements with the search model used. Therefore, the crystal structure of *E. coli* common pilus A (EcpA), which also lacks disulphide bonds, was used but molecular replacement remained unsuccessful.

Space Group	Translational Z-score	Packing Check (PAK)	Log Likelihood Gain
P6 ₅ 22	4.4	7	19
P6 ₁ 22	4.5	0	17
	4.8	0	16
	4.7	0	16
P6 ₃ 22	4.4	3	23

Table 8.3: Z-score of the partial solutions found from PHASER. Both Z-score and log likelihood are low and packing check indicates clashes in some cases.

8.3 Alternative Crystallization Conditions

Attempts to optimize the crystallisation conditions and also to implement macro-seeding, micro-seeding and additive screens were not sufficient to reproduce diffracting crystals. However, it was suggested that the use of D₂O instead of H₂O in both crystallisation and protein sample buffers may improve crystal formation as protein samples in D₂O display higher stability (Makhatadze *et al.* 1995; Sheu *et al.* 2008). It was suggested that the deuterated samples have strengthened hydrogen bonds and the changes in bond length hence may alter the geometry of hydrogen bond networks between molecules in a crystal (Loh and Markley 1994; Sigala *et al.* 2009). Therefore, a screen of different concentrations of D₂O was used to test this effect and its potential to improve the quality of protein crystals.

8.3.1 D₂O Crystallisation Conditions

The previous dataset was obtained from a crystallisation condition containing sodium acetate and ammonium acetate, which naturally became the first candidate for the screens of D₂O crystallisation (Table 8.4). Three sets of conditions were used to compare the effect of D₂O: 0%, 50% and 100% D₂O. Buffers were made in the same fashion with an additional set of buffers which used 100% D₂O to dissolve all chemicals. The 0% and 100% final crystallisation conditions were made up with only normal buffers (H₂O) and 100% D₂O buffers, respectively. The 50% final D₂O concentration was made up by 100% D₂O crystallisation buffer with 0% D₂O protein buffer.

Final percentage of D ₂ O	Protein	Buffer pH	Salt	Concentration of the Precipitant
0%	SefD	pH 4.0-5.4 of 0.1 M Sodium Acetate	-	3.7 M-4.2 M Ammonium Acetate
50%	SefD	pH 4.0-5.4 of 0.1 M D ₂ O Sodium Acetate	-	3.7 M-4.2 M D ₂ O Ammonium Acetate
100%	SefD in D ₂ O buffer	pH 4.0-5.4 of 0.1 M D ₂ O Sodium Acetate	-	3.7 M-4.2 M D ₂ O Ammonium Acetate

Table 8.4: Conditions for D₂O crystallisation screens. The conditions remained the same but D₂O was used to make up either the crystallisation buffer or the protein sample buffer to achieve different percentage of D₂O in the final crystallisation condition.

8.3.2 Crystal Formation in D₂O

All samples were incubated at 20 °C and monitored after a few days. The 0% D₂O samples showed either clear drops or precipitation. In the 50% D₂O samples some crystals were observed and noticeably many spherulite crystals (mushroom-like) were present which demonstrated the beginning of crystal formation. Furthermore, well-formed crystals were observed in 100% D₂O with approximate dimensions of 150x50x50 µm (Fig. 8.3).

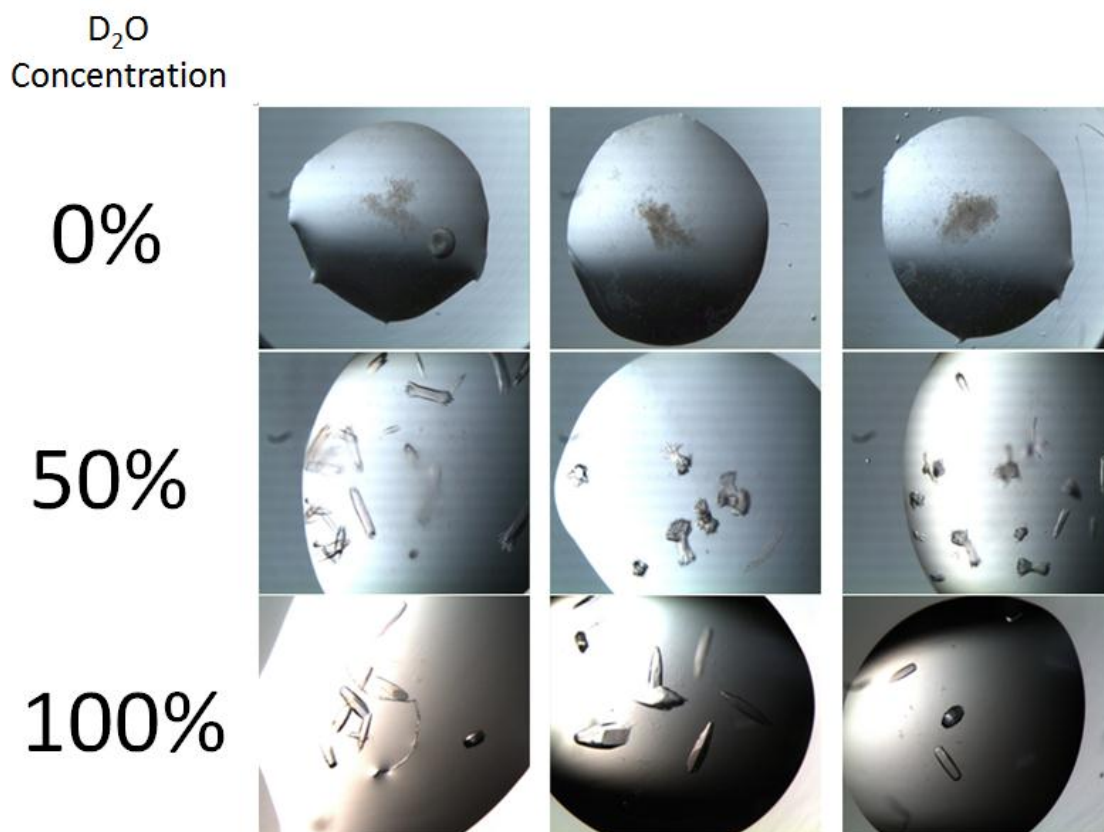


Figure 8.3: Images of crystals grown in different concentrations of D_2O . Precipitation and clear drops were observed in 0% D_2O samples, whereas 50% D_2O samples showed evidence of crystal growth. The presence of crystals was easily noticed in 100% D_2O samples.

8.4 Crystals from D_2O Conditions

Due to the unsuccessful molecular replacement with previous datasets, crystals were soaked in compounds amenable to anomalous dispersion phasing experiments namely, sodium iodide and 5-amino-2, 4, 6-triiodoisophthalic acid (I3C).

8.4.1 Data Collection

The crystals from D_2O conditions were soaked in either the same cryoprotectant as before or with the additional of 2 M sodium iodide or 0.25 M I3C for 60 seconds and flash frozen in liquid nitrogen and taken to beamline I24 (Diamond Laboratories, UK) where they were exposed to synchrotron radiation with wavelength 1.54980 Å. The

crystals gave similar diffraction of 3.0 Å. The point group was suggested to be P222 using two images separated by 90°, although to obtain highly redundant data, 360° and 720° of data were collected for native and soaked crystals, respectively, at 100 °K with oscillation angle of 0.2° per frame to 3.0 Å on a Pilatus 6M detector, with a detector distance of 527.77 mm.

8.4.2 Data Processing

8.4.2.1 Native Dataset

It was not possible to index the new data in either point groups of P6, P622 or C222 as suggested from the previous dataset but instead point group P222 was suggested. The Patterson self-rotation function indicated a strong 2-fold axis parallel to the *a*, *b* and *c* axes with weaker 3- and 6-fold symmetry along the crystallographic *c*-axis (Fig. 8.4). However, as the data could not be processed into 6-fold symmetry point groups, this higher symmetry indicates a 6-fold non-crystallographic relationship within the asymmetric unit. The data was further examined with a native Patterson map which revealed the presence of pseudo-translational non-crystallographic symmetry (PST) related by the vector = 0.5, 0.5, 0.032 which was 38% the size of the origin peak (Fig. 8.5).

The datasets were processed to 3.0 Å using MOSFLM (1994) and scaled to 3.0 Å with SCALA (1994) (Table 8.5). Matthews coefficients suggested many possibilities ranging from one molecule to 11 molecules in the asymmetric unit (Table 8.6). However, based on the 0% D₂O data, a solvent content of 48.97% (6 molecules in asymmetric unit) is most likely. After re-indexing in all alternative space groups of P222, molecular replacement was performed using the models described in section

8.2 with MOLREP using the pseudo-translation function. PST can cause many problems during molecular replacement, especially during the translation search, and further due to inconsistencies with systematic absences, the true space group is not always obvious until the model has been substantially refined. Therefore, molecular replacement was unsuccessful and the true space group is still unclear.

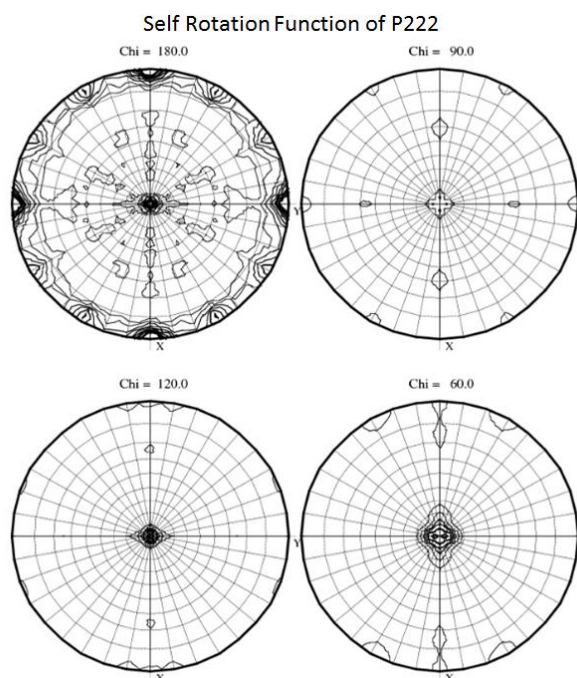


Figure 8.4: Patterson Self-Rotation map of SefD indexed in space group P222. It demonstrates strong 2-fold peaks in $\text{Chi}=180^\circ$ along the a , b and c -axis and a possible 3 and 6 fold NCS symmetry parallel to the crystallographic c -axis, as broad peaks observed in $\text{Chi}=120^\circ$ and 60° .

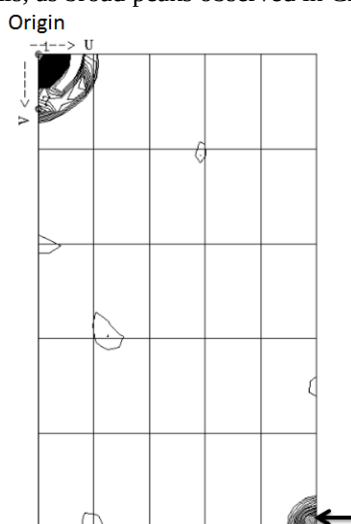


Figure 8.5: Native Patterson map of P222. The arrow indicates the vector of 38% height of the origin of a molecule present at $u = 0.5$, $v = 0.5$ and $w = 0.032$ caused by non-crystallographic pseudo-translation.

Protein		SefD (with SefA DSC strand)
Point Group		P222
Cell dimensions (Å)	a	52.59
	b	88.14
	c	212.03
Cell angles (°)	α	90.00
	β	90.00
	γ	90.00
Number of reflections (total)		91686
Number of reflections (unique)		16930
I/sigI (mean)		4.4 (1.7)
Multiplicity		5.4 (5.3)
R _{sym} (%)		12 (40)
Completeness (%)		99.5 (99.6)
Resolution range (Å)		88.14 – 3.2 (3.37 – 3.20)
Average Wilson B Value (Å ²)		83.047
Ranom		0.082 (0.06)
Rpim		0.084 (0.282)
RCR_anom		1.582

Table 8.5: Summary of data collection and scaling statics for SefD. The data is processed in the point group of P222. The statistics for the outliershell are in bracket.

Number of molecules per Asymmetric unit	Matthews Coefficient (Vm)	Percentage of Solvent Content
1	14.45	91.50
2	7.23	82.99
3	4.82	74.49
4	3.61	65.98
5	2.89	57.48
6	2.41	48.97
7	2.06	40.48
8	1.81	31.96
9	1.61	23.46
10	1.45	14.95
11	1.31	6.45

Table 8.6: Matthews Coefficient of SefD in D₂O. The number of molecules per asymmetric unit is given with the corresponding Matthews coefficient score and solvent content.

8.4.2.2 Sodium Iodide and I3C Soaks

The I3C data did not contain enough signals to calculate phase information, there was significant incorporation of iodide to have a clear signal out to around 4 Å. However, likely due to the presence of many iodide ions bound with low occupancy, it was not possible to solve the structure using Single Anomalous Dispersion (SAD) techniques.

It may still be possible to overcome this by repeating these experiments using Multiple Anomalous Dispersion (MAD) (Table 8.5).

8.5 Selenomethionine Samples

Substitution of methionine to selenomethionine could allow anomalous scattering of selenium to be recorded for phase information. Therefore, *E. coli* cells were grown in minimum media as described in Material and Methods (Chapter 3), but instead selenomethionine was added to the medium to allow production of selenomethionine protein samples. These protein samples were used for crystallisation with normal buffer (0% D₂O) and 100% D₂O buffer as described earlier. Precipitation and clear drops were observed in the 0% D₂O condition, but similar crystals were harvested from 100% D₂O conditions.

A fluorescence scan indicated there were not enough selenomethionine signals due to low incorporation in the protein. As the fact there are only two methionines in the protein sequence whilst one of them is located near the N-terminus, therefore low signal of selenomethionine was anticipated.

8.6 Summary

The initial datasets of SefD were collected and indicated the point group of P622 and C222. As the self-rotation function suggested strong 2- and 6-fold symmetry, the likely true point group is P622. However, molecular replacement was unsuccessful with all the search models in all possible space group derivatives. Therefore, other phasing techniques are needed for structure determination. Unfortunately, the crystals are not reproducible by standard approaches.

New crystals were obtained from the same crystallisation condition but with D₂O solutions instead. Rather than solely depending on molecular replacement, the crystals were also soaked in sodium iodide and I3C for performing other phasing techniques. Furthermore, selenomethionine samples of SefD were also produced and crystallized in D₂O condition. However, these new derivative crystals demonstrated no signal for selenomethionine, low signal for the I3C soaks and multiple low occupancy incorporation for sodium iodide. Therefore, it has not been possible to resolve the phase problem. Molecular replacement remained unsuccessful with the new dataset but the data suggested a different point group, P222.

8.6.1 Effect of D₂O

Since the structure of the 0% D₂O crystals has not been solved both point groups, P622 and C222, are still candidates. However, space group C222 is a subgroup of, and mathematically related to the higher symmetry space group P622. The 100% D₂O datasets suggest a point group of P222 which is similarly related to C222 by the pseudo-translational non-crystallographic symmetry vector 0.5, 0.5, 0.032. The cell dimensions of P222 and C222 are very similar (Table 8.7) and together these indicate that they form a very similar lattice. In point group P222, there are four asymmetric units present in the unit cell and point group C222 contains eight asymmetric units (Fig. 8.6).

Although this theory on the different point groups will need to be confirmed once the structure is determined, in the case of SefD, D₂O has proved to facilitate the crystallisation.

Point Group		P222	C222
Cell dimensions (Å)	a	52.29	52.2585
	b	88.1399	90.47
	b	212.0299	216.8473
Cell angles (°)	α	90.00	90.00
	β	90.00	90.00
	γ	90.00	90.00

Table 8.7: Cell dimensions and angles of P222 and C222.

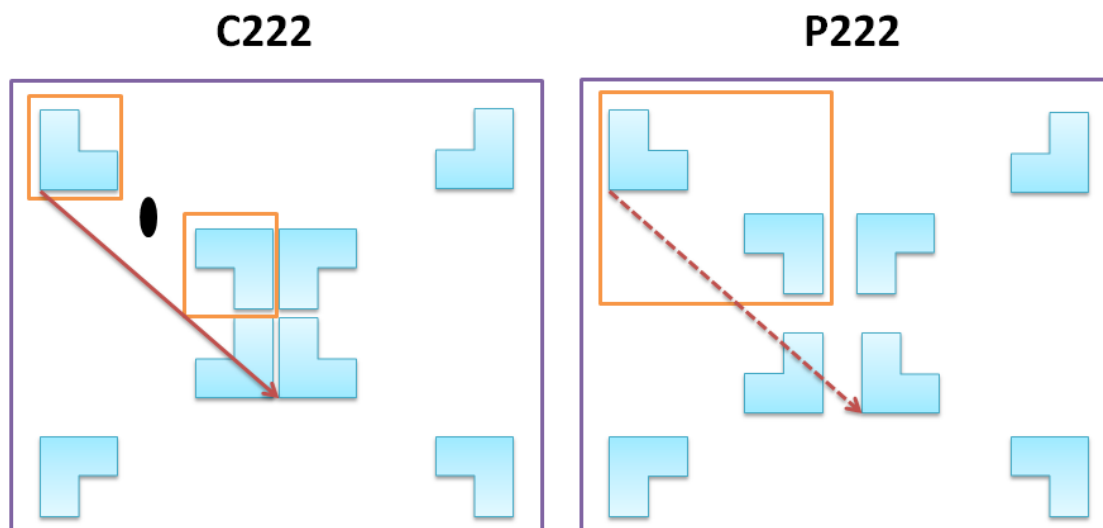


Figure 8.6: Asymmetric units in unit cell of C222 and P222. Each molecule is represented by a blue L-shaped block. The purple lines are the boundaries of the unit cell and a typical asymmetric unit is in the orange square. In C222, each molecule is crystallographically related and there are 8 asymmetric units in the unit cell. As for P222 the asymmetric unit contains two molecules and the total of 4 asymmetric units make up the unit cell. The red solid and dash arrow indicates crystallographic symmetry and non-crystallographic pseudo-translation symmetry.

Chapter 9: Conclusions and Future Perspectives

9.1 Structural Studies of AafB and HdaB

EAEC is one of the six pathogenic *E. coli* that causes intestinal and urinary tract infection and is responsible for many outbreaks worldwide including the recent hybrid strain affecting Europe (Bielaszewska *et al.* 2011; Frank *et al.* 2011; Scheutz *et al.* 2011; Struelens *et al.* 2011). It has the characteristic aggregative adherence properties achieved through surface fimbrial proteins. The fimbriae can be divided into 4 classes, Aaf-I (Agg-I), Aaf-II, Aaf-III (Agg-III) and Hda. These Aaf proteins share a similar chaperone-usher system to the Afa/Dr family. AafB (57%) and HdaB (45%) have high sequence identity to AfaD from Afa-III which acts as an invasin. Based on the understanding of the donor exchange mechanism of these fimbrial proteins, both AafB and HdaB were successfully cloned and expressed.

9.1.1 Structure Determination of AafB and HdaB

X-ray crystallography was used to obtain the structure of AafB. The structure has a resolution of 3.0 Å and the phase information was acquired by molecular replacement using DraD as the search model. The asymmetric unit contains two molecules and the unit cell is comprised of 12 asymmetric units with the space group of P6₁22. The structure of AafB demonstrates an immunoglobulin-like fold of a β -sandwich which consists of 9 β -strands, including the engineered donor strand on the C-terminus.

X-ray crystallography was also used to obtain a 2.0 Å resolution structure of HdaB. Molecular replacement using AafB as the search model was used to obtain phase information. There are two molecules in the asymmetric unit, and eight asymmetrical units are included in the unit cell. Despite the small differences between the two molecules in the asymmetric unit, the structure of HdaB displays a similar immunoglobulin-like fold to AafB with 9 β -strands including the engineered donor

strand on the C-terminus. Furthermore, the additional cysteine designed on the donor strand promotes the formation of a dimer which facilitates the stability of the protein and may have contributed to the high resolution data.

9.1.2 Structural Comparisons of AafB and HdaB to Afa/Dr homologues

Both AafB and HdaB display a high degree of sequence similarity and similar secondary structure to the Afa/Dr family. The structures demonstrate a common immunoglobulin-like fold with 9 β -strands including the additional engineered donor strand. All β -strands share a similar position and orientation in the tertiary structure as many conserved residues are present in these regions.

The main differences are located in the loop regions and the designed donor strand on the C-terminus. The loops are less conserved and variations of these loops are commonly found in other homologous proteins. The differences in the donor strand are caused by the different residues and length used in the clones.

9.2 Functional Studies of AafB and HdaB

9.2.1 Potential Binding Sites of AafB

Electrostatic analysis of AafB indicated many negatively- and positively-charged surface regions. A few ions and waters were found in the crystal structure of AafB, in particular a sulphate ion was observed in the centre of the positively-charged surface. This interaction might indicate a potential binding site for negatively-charged ligands.

9.2.2 NMR studies of AafB

An oligosaccharide screen was performed by our collaborators and exhibited positive interaction between AafB and the family of GAGs. In particular, chondroitin sulphate and heparin sulphate demonstrated the tightest binding.

These interaction studies were further pursued by NMR experiments. Despite an unexplained “doubling” effect of AafB, backbone assignments were achieved to around 65%. The missing loop regions of the AafB crystal structure remained absent in the NMR assignments and indicate a high degree of flexibility of these regions. In addition, the missing residues from the NMR experiments were identified as loop or flexible β -strand/loop similar to the loop regions of the AfaD NMR structure (Cota *et al.* 2006).

9.2.3 Interaction Studies of AafB

Several peaks shifted when AafB was covalently linked with AafA, and some residues were identified which were located near the donor strand and lower region of AafB. As the interface (lower region) between AafB and AafA is close in space, changes in the chemical environment causing peak shifts might be expected.

In addition, AafB was also titrated with chondroitin derivatives which resulted in heavy precipitation. The presence of precipitation might suggest some kind of interaction but the remaining soluble protein did not show any changes in the HSQC spectra. Nevertheless, small peak shifts were observed when AafB was titrated with heparin I-S (pH5). Three residues were identified; Gly 116, His 37 and Thr 24. Histidine is positively charged (at pH 5.4) and may be involved in direct interactions with the negatively charged sulphate groups within heparin. Changes in the chemical

shifts of the glycine and threonine residue might be the result of changes in the adjacent β -strand on binding or the formation another binding site.

9.2.4 Potential Binding site of HdaB

Three citrate ions have been observed interacting with positively-charged areas of HdaB which might suggest the location of potential binding sites. In addition, Thr 24 and Gly 116 of AafB are conserved in HdaB (as well as several residues nearby); this suggests a potential conserved binding site. Alternatively, these residues may be the result of conformational changes rather than a direct interaction. The significance of these changes needs further verification. The positively-charged residue, Arg 36 is located at the same position as His 37 of AafB which suggests a potential binding site for ligands.

9.3 Effect of D₂O in SefD Structural Studies

The initial datasets of SefD suggested a point group of P622 or C222 but as the phases could not be solved, the structure was not determined. Despite SefD crystals not being reproducible with common approaches, the use of D₂O yielded new crystal forms. The new datasets suggested a different point group, P222, which is related to C222 by non-crystallographic pseudo-translational symmetry. Molecular replacement remained unsuccessful with the new datasets and neither the iodide, I3C or selenomethionine derivatives provided sufficient information for phasing. The differences in the point group suggested the effect of D₂O might have altered the hydrogen bond network in the molecules. The effect of D₂O would need further studies to confirm.

9.4 Future Perspectives

9.4.1 Interaction Studies of AafB

Due to the appearance of doubled peaks in HSQC spectra of AafB, some peaks that changed position on addition of ligand were unidentified. However, co-crystallisation of an AafB AafA fusion with ligands might provide insight into the interaction. The initial screens for co-crystallisation of AafB and AafA were not successful but this might be caused by inter-domain flexibility. As the oligosaccharide screens suggested significant binding of the ligand by the two-domain construct, the ligand might be able to “lock” the domains, so co-crystallisation of ligands with the complex might yield some workable crystals.

Furthermore, the interaction of AafB and AafBA observed in NMR spectra was much weaker than the microarray experiments. One of the main factor could be the length of the ligands used in the experiments were different. The heparins used in NMR titration were a dimer but in microarray screens long polymers (6 to 12) were used. Therefore, long polymers should be included in further titration.

9.4.2 Functional Studies of HdaB

Oligosaccharide screens with HdaB would be ideal to identify potential binding partners and the stable dimer of HdaB might also be suitable for titration experiments. The fusion of HdaA and HdaB did not crystallise from the initial screens which might be caused by interdomain flexibility as well, but co-crystallisation with ligand might restrict the movement of these domains and hence should be considered.

9.4.3 Structural Studies of SefD

Screens will be required to identify the appropriate heavy metals to incorporate into SefD crystals for acquiring phase information. Once the structure of SefD is determined it will be possible to investigate the atomic details which have led to a change of point group caused by D₂O. Furthermore, other protein systems should be tested to confirm whether the effects of D₂O on improving the crystallisation conditions and resolution are a general trait of this solvent.

Appendices

Appendix 1: Sequence Alignment of AfaD-III homologues

```

AfaD-III      -----MNGSIRKMMRVTCGMLLMVMSG-----VSQAELHLES-RGGSGTQLRDGAKV 47
AfaD-I        -----MNGSIRKMMRVTCGMLLMVMSG-----VSQAELHLES-RGGSGTQLRNGAKL 47
AfaD-V        -----MNGSIRKMMRVTCGMLLMVMSG-----VSQAELTLES-HGSRGGQLRDGEKV 47
AfaD-II       -----MNGSIRKMMRVTCGMLLMVMSG-----VSQAELTLES-HGSRGGQLRDGEKL 47
AggB          -----MLK---KSILPMSCGVLVMVMSG-----LLDAAELITLIS-HKTLGSQLRDGMKL 45
HdaB          -----MK---LSISLTLCGILLMLMGS-----FSQAADITLMN-HKYMGNLLHDGVKL 44
Agg3B         -----MMK---KGIVSLSCGACLMVVS-----LSAAAEINLVS-HRNTESQLRDGMKL 45
AafB          -----MK---KGMLSVSCGMLLMVISG-----LSQATEISLEGLHRNMGQLFDGDIL 45
AfaD-VII      -----MKKNGKTHVLHGCLAVVMVMAG-----VSQAELSLDV-RKAMGSELRDGERI 48
AfaD-VIII     -----MK---KIQIVCSGIVLVVSS-----LAQAVELSLNT-SDGRSGELKDGTKV 43
SafD          MWMKIQRVKTVIYSVSLVAASSLPVIAN-----AAEKLTTLRVGTYFRAGHVPDGMVL 55
SefD          -----MNQYNSSIPKFIVSVFLIVTGFFSSTIKAQELKLMIKINEAVFYDRITSNKII 53
              ::                ::                .      :                : ..      :

AfaD-III      ATGRIICREAHTGFHVWMNERQVDGRAERYVVQSKDG-RHELVRVTGGDGWSPVKGEGGK 106
AfaD-I        ATGRIICREAHTGFHVWMNERQVDGRAERYVVQSKDG-RHELVRVTGGDGWSPVKGEGGK 106
AfaD-V        ATGRIICREAHTGFHVWMNERQVDGRAERYVVQSKDG-RHELVRVTGGDGWSPVKGEGGK 106
AfaD-II       ATGRIICREAHTGFHVWMNERQVDGRAERYVVQSKDG-RHELVRVTGGDGWSPVKGEGGK 106
AggB          ATGRIACREPHDGFHIWNASQNGK-VGHYIVQNNRETKEHLVKVIGGGWSSSLIEGQR 104
HdaB          ATGRIICQDTHSGFRVWINARQEGGGAGKYIVQSTEGPQHNLRIIGNGWSSFVEKGIQ 104
Agg3B         ATGRMTCRETHTGfHVWMNARQDGGRPHEYIVQNSKGIQHELVRIRIGNGWISSFGEAQR 105
AafB          ATGRIICREHRTGFHIQMNAHQVEGRPGHYIVQGSKDTQSKLWVRLGREGWTSPTEGGGQQ 105
AfaD-VII      ATGRIICREAHTGFHVWMNGREDEGRPGHLLIQKHDSRNEIRARVEGEWWSAIT-EGEH 107
AfaD-VIII     ATGRIICRGTYTSFHIWMNSRQMGNIPGHYIILGRHDSHNEMRVRLDAGWLPVSDGQ- 102
SafD          AQGWVITYHGSHSGFRVWSDQKAGNTPTVLLLSGQQDPRHHIQVRLGEQWQPDTVSGRG 115
SefD          GTGHLFNRREGKK-ILISSSLEKIKNTPGAYIIRGQNNSAHKLRIKIGGEDWQPDNSGIG- 111
              . * : : : : . : : : : : . : : : . * .

AfaD-III      GVSRLPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVVPQE-- 147
AfaD-I        GVSRLPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVVPQE-- 147
AfaD-V        GVSRLPGQEEQVFFDVMADGNQDIAPGEYRFSVGGACVGPQE-- 147
AfaD-II       GVSRLPGQEEQVFFDVMVDGNQNIAPGEYRFSVGGACVGPQE-- 147
AggB          GVSRLPGQEEQVFFDVMVDGNQNIAPGEYRFSVGGACVGPQE-- 147
HdaB          GVFNTIKEDASIFYIEVDGNQQVHPGKYLFSVSGECYI----- 142
Agg3B         GIFRLGKEEQAIQFDVIVDGDQKVIPGEYMLISGECIVLGR-- 146
AafB          GIVRSQGEQVIFDVMADGNQWAKPGEYIFSVSGKCLTSWE-- 146
AfaD-VII      RLVKAGEEQQAVFDVVS DGRQAGADEYVLSVSGSCVDSRSQE 150
AfaD-VIII     GMVSTGIPEQHTFDVVIDGNQLLGPDEYILSVSGECS----- 139
SafD          AILRTAADN-ASFVVDGNQEVPA DTWTLDFKACALAQEDT- 156
SefD          --MVSHSDFTNFNIYFFGNGDIPVD TYLIS IYATEIEL---- 148
              * : * . : : . .

```

Appendices 2: Primer sequences for building different constructs

Primer	DNA Sequence
afaDVII-F	5'- CACGGATCCGCGGAACTGAGCCTG - 3'
afaDVII-R	5'- GAGAAGCTTGCGAATAATAAAGCTGGTATC - 3'
afaEVII-F	5'- CGCGGATCCGATGATTGCGCGATTAG - 3'
afaEVII-R	5'- CTAATTAAGCTTGCGAATAATAAAGCTG - 3'
agg3B-F	5'- CACGGATCCGCTGAAATTAATC - 3'
agg3B-R	5'- GAGAAGCTTAACAAATTTTCAGTTTAGCAG - 3'
agg3A-F	5'- GTTGGATCCAATAAATGTTCTCTG - 3'
agg3A-R	5'- CTAATTAAGCTTAACAAATTTTC - 3'
hdaB-F	5'- CACGGATCCGCGGATATTAC - 3'
hdaB-R	5'- GAGAAGCTTTTCCACGGTATGCTGCGCGGTAATG - 3'
hdaA-F	5'- GAAGGATCCGCGGATGTGGAATTTAC - 3'
hdaA-R	5'- CTAATTAAGCTTTTCCACGGTATG - 3'
sefD-F	5'- CACGGATCCCTGAAACTGATGATTAATAAAC - 3'
sefD-R	5'- GAGAAGCTTCGCCGCAATGGTCACCGCCGCCTGCAC - 3'
sefA-F	5'- GCGGGATCCCAGAACACCACCAGCGCGAACTGGAG - 3'
sefA-R	5'- CTAATTAAGCTTCGCCGCAATGGTC - 3'

References

- (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-763.
- Adachi, J. A., Z. D. Jiang, et al. (2001). "Enterotoxigenic Escherichia coli as a major etiologic agent in traveler's diarrhea in 3 regions of the world." Clin Infect Dis **32**(12): 1706-1709.
- Adachi, J. A., J. J. Mathewson, et al. (2002). "Enteric pathogens in Mexican sauces of popular restaurants in Guadalajara, Mexico, and Houston, Texas." Ann Intern Med **136**(12): 884-887.
- Altomare, A., C. Cuocci, et al. (2008). "Minimally resolution biased electron-density maps." Acta Crystallogr A **64**(Pt 2): 326-336.
- Anderson, G. G., K. W. Dodson, et al. (2004). "Intracellular bacterial communities of uropathogenic Escherichia coli in urinary tract pathogenesis." Trends Microbiol **12**(9): 424-430.
- Anderson, K. L., J. Billington, et al. (2004). "An atomic resolution model for assembly, architecture, and function of the Dr adhesins." Mol Cell **15**(4): 647-657.
- Archambaud, M., P. Courcoux, et al. (1988). "Detection by molecular hybridization of pap, afa, and sfa adherence systems in Escherichia coli strains associated with urinary and enteral infections." Ann Inst Pasteur Microbiol **139**(5): 575-588.
- Barnhart, M. M., J. S. Pinkner, et al. (2000). "PapD-like chaperones provide the missing information for folding of pilin proteins." Proc Natl Acad Sci U S A **97**(14): 7709-7714.
- Baumler, A. J., R. M. Tsois, et al. (1996). "The pef fimbrial operon of Salmonella typhimurium mediates adhesion to murine small intestine and is necessary for fluid accumulation in the infant mouse." Infect Immun **64**(1): 61-68.
- Baumler, A. J., R. M. Tsois, et al. (1996). "The lpf fimbrial operon mediates adhesion of Salmonella typhimurium to murine Peyer's patches." Proc Natl Acad Sci U S A **93**(1): 279-283.
- Beachey, E. H. (1981). "Bacterial adherence: adhesin-receptor interactions mediating the attachment of bacteria to mucosal surface." J Infect Dis **143**(3): 325-345.
- Beaulieu, J. F. (1999). "Integrins and human intestinal cell functions." Front Biosci **4**: D310-321.
- Beinke, C., S. Laarmann, et al. (1998). "Diffusely adhering Escherichia coli strains induce attaching and effacing phenotypes and secrete homologs of Esp proteins." Infect Immun **66**(2): 528-539.
- Benz, I. and M. A. Schmidt (1989). "Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic Escherichia coli." Infect Immun **57**(5): 1506-1511.
- Benz, I. and M. A. Schmidt (1992). "AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is synthesized via a precursor molecule." Mol Microbiol **6**(11): 1539-1546.
- Benz, I. and M. A. Schmidt (1992). "Isolation and serologic characterization of AIDA-I, the adhesin mediating the diffuse adherence phenotype of the diarrhea-associated Escherichia coli strain 2787 (O126:H27)." Infect Immun **60**(1): 13-18.

- Benz, I. and M. A. Schmidt (1993). "Diffuse adherence of enteropathogenic Escherichia coli strains--processing of AIDA-I." Zentralbl Bakteriol **278**(2-3): 197-208.
- Bernier, C., P. Gounon, et al. (2002). "Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative Escherichia coli as a sensitive probe for detecting the AAF-encoding operon family." Infect Immun **70**(8): 4302-4311.
- Bhan, M. K., P. Raj, et al. (1989). "Enteroaggregative Escherichia coli associated with persistent diarrhea in a cohort of rural children in India." J Infect Dis **159**(6): 1061-1064.
- Bielaszewska, M., A. Mellmann, et al. (2011). "Characterisation of the Escherichia coli strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study." Lancet Infect Dis.
- Bilge, S. S., C. R. Clausen, et al. (1989). "Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated Escherichia coli to HEp-2 cells." J Bacteriol **171**(8): 4281-4289.
- Blanc-Potard, A. B., C. Tinsley, et al. (2002). "Representational difference analysis between Afa/Dr diffusely adhering Escherichia coli and nonpathogenic E. coli K-12." Infect Immun **70**(10): 5503-5511.
- Blow, D. M. (2002). Outline of crystallography for biologists. Oxford ; New York, Oxford University Press.
- Blundell, T. L. and L. N. Johnson (1976). Protein crystallography. New York, Academic Press.
- Boisen, N., C. Struve, et al. (2008). "New adhesin of enteroaggregative Escherichia coli related to the Afa/Dr/AAF family." Infect Immun **76**(7): 3281-3292.
- Brooks, J. T., E. G. Sowers, et al. (2005). "Non-O157 Shiga toxin-producing Escherichia coli infections in the United States, 1983-2002." J Infect Dis **192**(8): 1422-1429.
- Bullitt, E., C. H. Jones, et al. (1996). "Development of pilus organelle subassemblies in vitro depends on chaperone uncapping of a beta zipper." Proc Natl Acad Sci U S A **93**(23): 12890-12895.
- Cabeen, M. T. and C. Jacobs-Wagner (2005). "Bacterial cell shape." Nat Rev Microbiol **3**(8): 601-610.
- Capitani, G., O. Eidam, et al. (2006). "Evidence for a novel domain of bacterial outer membrane ushers." Proteins **65**(4): 816-823.
- Caprioli, A., S. Morabito, et al. (2005). "Enterohaemorrhagic Escherichia coli: emerging issues on virulence and modes of transmission." Vet Res **36**(3): 289-311.
- Cavanagh, J. (2007). Protein NMR spectroscopy : principles and practice. Amsterdam ; Boston, Academic Press.
- Chan, K. N., A. D. Phillips, et al. (1994). "Enteroaggregative Escherichia coli: another cause of acute and chronic diarrhoea in England?" J Pediatr Gastroenterol Nutr **18**(1): 87-91.
- Chen, T. and E. C. Gotschlich (1996). "CGM1a antigen of neutrophils, a receptor of gonococcal opacity proteins." Proc Natl Acad Sci U S A **93**(25): 14851-14856.
- Chen, V. B., W. B. Arendall, 3rd, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

- Choudhury, D., A. Thompson, et al. (1999). "X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*." Science **285**(5430): 1061-1066.
- Claridge, T. D. W. (2009). High-resolution NMR techniques in organic chemistry. Amsterdam ; Boston, Elsevier.
- Clore, G. M. and A. M. Gronenborn (1998). "NMR structure determination of proteins and protein complexes larger than 20 kDa." Curr Opin Chem Biol **2**(5): 564-570.
- Clouthier, S. C., S. K. Collinson, et al. (1994). "Unique fimbriae-like structures encoded by *sefD* of the SEF14 fimbrial gene cluster of *Salmonella enteritidis*." Mol Microbiol **12**(6): 893-901.
- Clouthier, S. C., K. H. Muller, et al. (1993). "Characterization of three fimbrial genes, *sefABC*, of *Salmonella enteritidis*." J Bacteriol **175**(9): 2523-2533.
- Collinson, S. K., L. Emody, et al. (1991). "Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis*." J Bacteriol **173**(15): 4773-4781.
- Cota, E., C. Jones, et al. (2006). "The solution structure of the invasive tip complex from Afa/Dr fibrils." Mol Microbiol **62**(2): 356-366.
- Cravioto, A., A. Tello, et al. (1991). "Association of *Escherichia coli* HEp-2 adherence patterns with type and duration of diarrhoea." Lancet **337**(8736): 262-264.
- Czczulin, J. R., S. Balepur, et al. (1997). "Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*." Infect Immun **65**(10): 4135-4145.
- D'Orazio, S. E. and C. M. Collins (1998). "Molecular pathogenesis of urinary tract infections." Curr Top Microbiol Immunol **225**: 137-164.
- Daigle, F., J. Harel, et al. (1994). "Expression and detection of *pap*-, *sfa*-, and *afa*-encoded fimbrial adhesin systems among uropathogenic *Escherichia coli*." Can J Microbiol **40**(4): 286-291.
- Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.
- Dibb-Fuller, M. P., E. Allen-Vercoe, et al. (1999). "Fimbriae- and flagella-mediated association with and invasion of cultured epithelial cells by *Salmonella enteritidis*." Microbiology **145** (Pt 5): 1023-1031.
- Dodson, K. W., F. Jacob-Dubuisson, et al. (1993). "Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes." Proc Natl Acad Sci U S A **90**(8): 3670-3674.
- Durrer, P., R. Zbinden, et al. (2000). "Intestinal infection due to enteroaggregative *Escherichia coli* among human immunodeficiency virus-infected persons." J Infect Dis **182**(5): 1540-1544.
- Edwards, R. A., D. M. Schifferli, et al. (2000). "A role for *Salmonella* fimbriae in intraperitoneal infections." Proc Natl Acad Sci U S A **97**(3): 1258-1262.
- Elias, W. P., Jr., J. R. Czczulin, et al. (1999). "Organization of biogenesis genes for aggregative adherence fimbria II defines a virulence gene cluster in enteroaggregative *Escherichia coli*." J Bacteriol **181**(6): 1779-1785.
- Emanuelsson, O., S. Brunak, et al. (2007). "Locating proteins in the cell using TargetP, SignalP and related tools." Nat Protoc **2**(4): 953-971.
- Falcao, J. P., D. P. Falcao, et al. (2004). "Ice as a vehicle for diarrheagenic *Escherichia coli*." Int J Food Microbiol **91**(1): 99-103.

- Farfan, M. J., K. G. Inman, et al. (2008). "The major pilin subunit of the AAF/II fimbriae from enteroaggregative *Escherichia coli* mediates binding to extracellular matrix proteins." Infect Immun **76**(10): 4378-4384.
- Feutrier, J., W. W. Kay, et al. (1986). "Purification and characterization of fimbriae from *Salmonella enteritidis*." J Bacteriol **168**(1): 221-227.
- Feutrier, J., W. W. Kay, et al. (1988). "Cloning and expression of a *Salmonella enteritidis* fimbrin gene in *Escherichia coli*." J Bacteriol **170**(9): 4216-4222.
- Frank, C., D. Werber, et al. (2011). "Epidemic Profile of Shiga-Toxin-Producing *Escherichia coli* O104:H4 Outbreak in Germany - Preliminary Report." N Engl J Med.
- Fritz, I., C. Strompl, et al. (2004). "Phylogenetic relationships of the genera *Stella*, *Labrys* and *Angulomicrobium* within the 'Alphaproteobacteria' and description of *Angulomicrobium amanitifforme* sp. nov." Int J Syst Evol Microbiol **54**(Pt 3): 651-657.
- Gallegos, M. T., R. Schleif, et al. (1997). "Arac/XylS family of transcriptional regulators." Microbiol Mol Biol Rev **61**(4): 393-410.
- Garcia, M. I., A. Labigne, et al. (1994). "Nucleotide sequence of the afimbrial-adhesin-encoding afa-3 gene cluster and its translocation via flanking IS1 insertion sequences." J Bacteriol **176**(24): 7601-7613.
- Gardner, K. H. and L. E. Kay (1998). "The use of ²H, ¹³C, ¹⁵N multidimensional NMR to study the structure and dynamics of proteins." Annu Rev Biophys Biomol Struct **27**: 357-406.
- Gioppo, N. M., W. P. Elias, Jr., et al. (2000). "Prevalence of HEp-2 cell-adherent *Escherichia coli* and characterisation of enteroaggregative *E. coli* and chain-like adherent *E. coli* isolated from children with and without diarrhoea, in Londrina, Brazil." FEMS Microbiol Lett **190**(2): 293-298.
- Goluszko, P., R. Selvarangan, et al. (1999). "Decay-accelerating factor and cytoskeleton redistribution pattern in HeLa cells infected with recombinant *Escherichia coli* strains expressing Dr family of adhesins." Infect Immun **67**(8): 3989-3997.
- Gonzalez, R., C. Diaz, et al. (1997). "Age-specific prevalence of *Escherichia coli* with localized and aggregative adherence in Venezuelan infants with acute diarrhea." J Clin Microbiol **35**(5): 1103-1107.
- Goodnough, M. C. and E. A. Johnson (1991). "Control of *Salmonella enteritidis* infections in poultry by polymyxin B and trimethoprim." Appl Environ Microbiol **57**(3): 785-788.
- Gray-Owen, S. D., D. R. Lorenzen, et al. (1997). "Differential Opa specificities for CD66 receptors influence tissue interactions and cellular response to *Neisseria gonorrhoeae*." Mol Microbiol **26**(5): 971-980.
- Guignot, J., M. F. Bernet-Camard, et al. (2001). "Polarized entry of uropathogenic Afa/Dr diffusely adhering *Escherichia coli* strain IH11128 into human epithelial cells: evidence for alpha5beta1 integrin recognition and subsequent internalization through a pathway involving caveolae and dynamic unstable microtubules." Infect Immun **69**(3): 1856-1868.
- Guignot, J., C. Chaplais, et al. (2007). "The secreted autotransporter toxin, Sat, functions as a virulence factor in Afa/Dr diffusely adhering *Escherichia coli* by promoting lesions in tight junction of polarized epithelial cells." Cell Microbiol **9**(1): 204-221.
- Harker, D. (1956). "X-ray diffraction applied to crystalline proteins." Adv Biol Med Phys **4**: 1-22.

- Huang, Y., B. S. Smith, et al. (2009). "Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC." Proc Natl Acad Sci U S A **106**(18): 7403-7407.
- Hultgren, S. J., F. Lindberg, et al. (1989). "The PapG adhesin of uropathogenic *Escherichia coli* contains separate regions for receptor binding and for the incorporation into the pilus." Proc Natl Acad Sci U S A **86**(12): 4357-4361.
- Hung, D. L., S. D. Knight, et al. (1996). "Molecular basis of two subfamilies of immunoglobulin-like chaperones." Embo J **15**(15): 3792-3805.
- Huppertz, H. I., S. Rutkowski, et al. (1997). "Acute and chronic diarrhoea and abdominal colic associated with enteroaggregative *Escherichia coli* in young children living in western Europe." Lancet **349**(9066): 1660-1662.
- Hwang, T. L. a. A. J. S. (1995). "Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients." Journal of Magnetic Resonance Series A **112**(2): 275-279.
- Jedrzejczak, R., Z. Dauter, et al. (2006). "Structure of DraD invasin from uropathogenic *Escherichia coli*: a dimer with swapped beta-tails." Acta Crystallogr D Biol Crystallogr **62**(Pt 2): 157-164.
- Johnson, J. R. (1991). "Virulence factors in *Escherichia coli* urinary tract infection." Clin Microbiol Rev **4**(1): 80-128.
- Jones, C. H., P. N. Danese, et al. (1997). "The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems." Embo J **16**(21): 6394-6406.
- Jouve, M., M. I. Garcia, et al. (1997). "Adhesion to and invasion of HeLa cells by pathogenic *Escherichia coli* carrying the afa-3 gene cluster are mediated by the AfaE and AfaD proteins, respectively." Infect Immun **65**(10): 4082-4089.
- Jung, Y. S. and M. Zweckstetter (2004). "Mars -- robust automatic backbone assignment of proteins." J Biomol NMR **30**(1): 11-23.
- Kang, G., M. M. Mathan, et al. (1995). "Evaluation of a simplified HEp-2 cell adherence assay for *Escherichia coli* isolated from south Indian children with acute diarrhea and controls." J Clin Microbiol **33**(8): 2204-2205.
- Kaper, J. B., J. P. Nataro, et al. (2004). "Pathogenic *Escherichia coli*." Nat Rev Microbiol **2**(2): 123-140.
- Keeler, J. (2005). Understanding NMR spectroscopy. Chichester, U.K., John Wiley and Sons.
- Keller, R., J. G. Ordóñez, et al. (2002). "Afa, a diffuse adherence fibrillar adhesin associated with enteropathogenic *Escherichia coli*." Infect Immun **70**(5): 2681-2689.
- Kendrew, J. C. (1962). "The structure of globular proteins." Comp Biochem Physiol **4**: 249-252.
- Knutton, S., R. K. Shaw, et al. (1992). "Ability of enteroaggregative *Escherichia coli* strains to adhere in vitro to human intestinal mucosa." Infect Immun **60**(5): 2083-2091.
- Kojima, S. and D. F. Blair (2004). "The bacterial flagellar motor: structure and function of a complex molecular machine." Int Rev Cytol **233**: 93-134.
- Kuehn, M. J., S. Normark, et al. (1991). "Immunoglobulin-like PapD chaperone caps and uncaps interactive surfaces of nascently translocated pilus subunits." Proc Natl Acad Sci U S A **88**(23): 10586-10590.
- Kuehn, M. J., D. J. Ogg, et al. (1993). "Structural basis of pilus subunit recognition by the PapD chaperone." Science **262**(5137): 1234-1241.

- Kumar, A., H. J. Chiu, et al. (2010). "Ligands in PSI structures." Acta Crystallogr Sect F Struct Biol Cryst Commun **66**(Pt 10): 1309-1316.
- Labigne-Roussel, A. F., D. Lark, et al. (1984). "Cloning and expression of an afimbrial adhesin (AFA-I) responsible for P blood group-independent, mannose-resistant hemagglutination from a pyelonephritic Escherichia coli strain." Infect Immun **46**(1): 251-259.
- Lalioui, L., M. Jouve, et al. (1999). "Molecular cloning and characterization of the afa-7 and afa-8 gene clusters encoding afimbrial adhesins in Escherichia coli strains associated with diarrhea or septicemia in calves." Infect Immun **67**(10): 5048-5059.
- Lalioui, L. and C. Le Bouguenec (2001). "afa-8 Gene cluster is carried by a pathogenicity island inserted into the tRNA(Phe) of human and bovine pathogenic Escherichia coli isolates." Infect Immun **69**(2): 937-948.
- Le Bouguenec, C., M. I. Garcia, et al. (1993). "Characterization of plasmid-borne afa-3 gene clusters encoding afimbrial adhesins expressed by Escherichia coli strains associated with intestinal or urinary tract infections." Infect Immun **61**(12): 5106-5114.
- Le Bouguenec, C. and A. L. Servin (2006). "Diffusely adherent Escherichia coli strains expressing Afa/Dr adhesins (Afa/Dr DAEC): hitherto unrecognized pathogens." FEMS Microbiol Lett **256**(2): 185-194.
- Li, H., L. Qian, et al. (2004). "The outer membrane usher forms a twin-pore secretion complex." J Mol Biol **344**(5): 1397-1407.
- Lindler, L. E. and B. D. Tall (1993). "Yersinia pestis pH 6 antigen forms fimbriae and is induced by intracellular association with macrophages." Mol Microbiol **8**(2): 311-324.
- Loh, S. N. and J. L. Markley (1994). "Hydrogen bonding in proteins as studied by amide hydrogen D/H fractionation factors: application to staphylococcal nuclease." Biochemistry **33**(4): 1029-1036.
- Louvard, D., M. Kedinger, et al. (1992). "The differentiating intestinal epithelial cell: establishment and maintenance of functions through interactions between cellular structures." Annu Rev Cell Biol **8**: 157-195.
- Makhatadze, G. I., G. M. Clore, et al. (1995). "Solvent isotope effects and protein stability." Nature Structural Biology **2**(10): 852-855.
- Manting, E. H. and A. J. Driessen (2000). "Escherichia coli translocase: the unravelling of a molecular machine." Mol Microbiol **37**(2): 226-238.
- Mathewson, J. J., R. A. Oberhelman, et al. (1987). "Enteroadherent Escherichia coli as a cause of diarrhea among children in Mexico." J Clin Microbiol **25**(10): 1917-1919.
- Mayer, H. B. and C. A. Wanke (1995). "Enteraggregative Escherichia coli as a possible cause of diarrhea in an HIV-infected patient." N Engl J Med **332**(4): 273-274.
- McDaniel, T. K., K. G. Jarvis, et al. (1995). "A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens." Proc Natl Acad Sci U S A **92**(5): 1664-1668.
- Miettinen, A., B. Westerlund, et al. (1993). "Binding of bacterial adhesins to rat glomerular mesangium in vivo." Kidney Int **43**(3): 592-600.
- Montelione, G. T., D. Zheng, et al. (2000). "Protein NMR spectroscopy in structural genomics." Nat Struct Biol **7** **Suppl**: 982-985.
- Mossoro, C., P. Glaziou, et al. (2002). "Chronic diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome associated with HEp-2 adherent Escherichia coli

- in adults infected with human immunodeficiency virus in Bangui, Central African Republic." J Clin Microbiol **40**(8): 3086-3088.
- Muller, K. H., S. K. Collinson, et al. (1991). "Type 1 fimbriae of *Salmonella enteritidis*." J Bacteriol **173**(15): 4765-4772.
- Nataro, J. P. and J. B. Kaper (1998). "Diarrheagenic *Escherichia coli*." Clin Microbiol Rev **11**(1): 142-201.
- Nataro, J. P., J. B. Kaper, et al. (1987). "Patterns of adherence of diarrheagenic *Escherichia coli* to HEp-2 cells." Pediatr Infect Dis J **6**(9): 829-831.
- Nataro, J. P., I. C. Scaletsky, et al. (1985). "Plasmid-mediated factors conferring diffuse and localized adherence of enteropathogenic *Escherichia coli*." Infect Immun **48**(2): 378-383.
- Nataro, J. P., T. Steiner, et al. (1998). "Enteraggative *Escherichia coli*." Emerg Infect Dis **4**(2): 251-261.
- Nataro, J. P., D. Yikang, et al. (1993). "Aggregative adherence fimbria I expression in enteraggative *Escherichia coli* requires two unlinked plasmid regions." Infect Immun **61**(3): 1126-1131.
- Nataro, J. P., D. Yikang, et al. (1994). "AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteraggative *Escherichia coli*." J Bacteriol **176**(15): 4691-4699.
- Natuzzi, E. S., P. C. Ursell, et al. (1993). "Nitric oxide synthase activity in the pregnant uterus decreases at parturition." Biochem Biophys Res Commun **194**(1): 1-8.
- Neuhaus D, W. M. (1989). The nuclear Overhauser effect in structural and conformational analysis., VCH, Weinheim ; Cambridge.
- Nishiyama, M., M. Vetsch, et al. (2003). "Identification and characterization of the chaperone-subunit complex-binding domain from the type 1 pilus assembly platform FimD." J Mol Biol **330**(3): 513-525.
- Nowicki, B., J. P. Barrish, et al. (1987). "Molecular cloning of the *Escherichia coli* O75X adhesin." Infect Immun **55**(12): 3168-3173.
- Nowicki, B., A. Hart, et al. (1993). "Short consensus repeat-3 domain of recombinant decay-accelerating factor is recognized by *Escherichia coli* recombinant Dr adhesin in a model of a cell-cell interaction." J Exp Med **178**(6): 2115-2121.
- Nowicki, B., A. Labigne, et al. (1990). "The Dr hemagglutinin, afimbrial adhesins AFA-I and AFA-III, and F1845 fimbriae of uropathogenic and diarrhea-associated *Escherichia coli* belong to a family of hemagglutinins with Dr receptor recognition." Infect Immun **58**(1): 279-281.
- Nowicki, B., J. Moulds, et al. (1988). "A hemagglutinin of uropathogenic *Escherichia coli* recognizes the Dr blood group antigen." Infect Immun **56**(5): 1057-1060.
- Nowicki, B., R. Selvarangan, et al. (2001). "Family of *Escherichia coli* Dr adhesins: decay-accelerating factor receptor recognition and invasiveness." J Infect Dis **183 Suppl 1**: S24-27.
- Ogunniyi, A. D., I. Kotlarski, et al. (1997). "Role of SefA subunit protein of SEF14 fimbriae in the pathogenesis of *Salmonella enterica* serovar Enteritidis." Infect Immun **65**(2): 708-717.
- Ogunniyi, A. D., P. A. Manning, et al. (1994). "A *Salmonella enteritidis* 11RX pilin induces strong T-lymphocyte responses." Infect Immun **62**(12): 5376-5383.
- Painter, J. and E. A. Merritt (2006). "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion." Acta Crystallogr D Biol Crystallogr **62**(Pt 4): 439-450.

- Patterson, A. L. (1934). "A Fourier series method for the determination of the components of interatomic distances in crystals." Physics Review **46**: 372-376.
- Paul, M., T. Tsukamoto, et al. (1994). "The significance of enteroaggregative *Escherichia coli* in the etiology of hospitalized diarrhoea in Calcutta, India and the demonstration of a new honey-combed pattern of aggregative adherence." FEMS Microbiol Lett **117**(3): 319-325.
- Pellecchia, M., P. Sebbel, et al. (1999). "Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR." Nat Struct Biol **6**(4): 336-339.
- Pham, T. Q., P. Goluszko, et al. (1997). "Molecular cloning and characterization of Dr-II, a nonfimbrial adhesin-I-like adhesin isolated from gestational pyelonephritis-associated *Escherichia coli* that binds to decay-accelerating factor." Infect Immun **65**(10): 4309-4318.
- Phan, G., H. Remaut, et al. (2011). "Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate." Nature **474**(7349): 49-53.
- Pitari, G. M., L. V. Zingman, et al. (2003). "Bacterial enterotoxins are associated with resistance to colon cancer." Proc Natl Acad Sci U S A **100**(5): 2695-2699.
- Pohl, P., P. Lintermans, et al. (1991). "Epidemiological study of *Salmonella enteritidis* strains of animal origin in Belgium." Epidemiol Infect **106**(1): 11-16.
- Polotsky, Y., J. P. Nataro, et al. (1997). "Adherence patterns of bacterial diarrheal agents in AIDS." Adv Exp Med Biol **412**: 367-371.
- Pupo, G. M., R. Lan, et al. (2000). "Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics." Proc Natl Acad Sci U S A **97**(19): 10567-10572.
- Ramplung, A., J. R. Anderson, et al. (1989). "*Salmonella enteritidis* phage type 4 infection of broiler chickens: a hazard to public health." Lancet **2**(8660): 436-438.
- Rashid, R. A., P. I. Tarr, et al. (2006). "Expression of the *Escherichia coli* IrgA homolog adhesin is regulated by the ferric uptake regulation protein." Microb Pathog **41**(6): 207-217.
- Remaut, H., R. J. Rose, et al. (2006). "Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism." Mol Cell **22**(6): 831-842.
- Rivera, M. J., N. Rivera, et al. (1991). "Molecular and epidemiological study of *Salmonella* clinical isolates." J Clin Microbiol **29**(5): 927-932.
- Robertson, J., M. Gomersall, et al. (1975). "Mycoplasma hominis: growth, reproduction, and isolation of small viable cells." J Bacteriol **124**(2): 1007-1018.
- Rowe, P. C., E. Orrbine, et al. (1993). "Diarrhoea in close contacts as a risk factor for childhood haemolytic uraemic syndrome. The CPKDRC co-investigators." Epidemiol Infect **110**(1): 9-16.
- Sansonetti, P. J., A. Phalipon, et al. (2000). "Caspase-1 activation of IL-1 β and IL-18 are essential for *Shigella flexneri*-induced inflammation." Immunity **12**(5): 581-590.
- Sattler, M. (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Progress in Nuclear Magnetic Resonance Spectroscopy **34**(2): 93-158.
- Sauer, F. G., M. Barnhart, et al. (2000). "Chaperone-assisted pilus assembly and bacterial attachment." Curr Opin Struct Biol **10**(5): 548-556.

- Sauer, F. G., K. Futterer, et al. (1999). "Structural basis of chaperone function and pilus biogenesis." *Science* **285**(5430): 1058-1061.
- Sauer, F. G., M. A. Mulvey, et al. (2000). "Bacterial pili: molecular mechanisms of pathogenesis." *Curr Opin Microbiol* **3**(1): 65-72.
- Sauer, F. G., J. S. Pinkner, et al. (2002). "Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation." *Cell* **111**(4): 543-551.
- Saulino, E. T., D. G. Thanassi, et al. (1998). "Ramifications of kinetic partitioning on usher-mediated pilus biogenesis." *Embo J* **17**(8): 2177-2185.
- Savarino, S. J., P. Fox, et al. (1994). "Identification and characterization of a gene cluster mediating enteroaggregative Escherichia coli aggregative adherence fimbria I biogenesis." *J Bacteriol* **176**(16): 4949-4957.
- Scaletsky, I. C., S. H. Fabbri, et al. (2002). "HEp-2-adherent Escherichia coli strains associated with acute infantile diarrhea, Sao Paulo, Brazil." *Emerg Infect Dis* **8**(8): 855-858.
- Scaletsky, I. C., M. L. Silva, et al. (1984). "Distinctive patterns of adherence of enteropathogenic Escherichia coli to HeLa cells." *Infect Immun* **45**(2): 534-536.
- Schagger, H. and G. von Jagow (1987). "Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa." *Anal Biochem* **166**(2): 368-379.
- Scheut, F., E. Moller Nielsen, et al. (2011). "Characteristics of the enteroaggregative Shiga toxin/verotoxin-producing Escherichia coli O104:H4 strain causing the outbreak of haemolytic uraemic syndrome in Germany, May to June 2011." *Euro Surveill* **16**(24).
- Schroder, G. F., M. Levitt, et al. (2010). "Super-resolution biomolecular crystallography with low-resolution data." *Nature* **464**(7292): 1218-1222.
- Selvarangan, R., P. Goluszko, et al. (2000). "Role of decay-accelerating factor domains and anchorage in internalization of Dr-fimbriated Escherichia coli." *Infect Immun* **68**(3): 1391-1399.
- Servin, A. L. (2005). "Pathogenesis of Afa/Dr diffusely adhering Escherichia coli." *Clin Microbiol Rev* **18**(2): 264-292.
- Sheikh, J., S. Hicks, et al. (2001). "Roles for Fis and YafK in biofilm formation by enteroaggregative Escherichia coli." *Mol Microbiol* **41**(5): 983-997.
- Sheldrick, G. M. (2008). "A short history of SHELX." *Acta Crystallogr A* **64**(Pt 1): 112-122.
- Sheu, S.-Y., E. W. Schlag, et al. (2008). "Molecular Dynamics of Hydrogen Bonds in Protein-D₂O: The Solvent Isotope Effect." *The Journal of Physical Chemistry A* **112**(5): 797-802.
- Sigala, P. A., M. A. Tsuchida, et al. (2009). "Hydrogen bond dynamics in the active site of photoactive yellow protein." *Proc Natl Acad Sci U S A* **106**(23): 9232-9237.
- Silverman, P. M. (1997). "Towards a structural biology of bacterial conjugation." *Mol Microbiol* **23**(3): 423-429.
- Slonim, L. N., J. S. Pinkner, et al. (1992). "Interactive surface in the PapD chaperone cleft is conserved in pilus chaperone superfamily and essential in subunit recognition and assembly." *Embo J* **11**(13): 4747-4756.
- So, S. S. and D. G. Thanassi (2006). "Analysis of the requirements for pilus biogenesis at the outer membrane usher and the function of the usher C-terminus." *Mol Microbiol* **60**(2): 364-375.

- Soto, G. E., K. W. Dodson, et al. (1998). "Periplasmic chaperone recognition motif of subunits mediates quaternary interactions in the pilus." Embo J **17**(21): 6155-6167.
- Soto, G. E. and S. J. Hultgren (1999). "Bacterial adhesins: common themes and variations in architecture and assembly." J Bacteriol **181**(4): 1059-1071.
- Spangler, B. D. (1992). "Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin." Microbiol Rev **56**(4): 622-647.
- Striker, R., F. Jacob-Dubuisson, et al. (1994). "Stable fiber-forming and nonfiber-forming chaperone-subunit complexes in pilus biogenesis." J Biol Chem **269**(16): 12233-12239.
- Struelens, M., D. Palm, et al. (2011). "Enterotoxigenic, Shiga toxin-producing *Escherichia coli* O104:H4 outbreak: new microbiological findings boost coordinated investigations by European public health laboratories." Euro Surveill **16**(24).
- Swanson, T. N., S. S. Bilge, et al. (1991). "Molecular structure of the Dr adhesin: nucleotide sequence and mapping of receptor-binding domain by use of fusion constructs." Infect Immun **59**(1): 261-268.
- Thanassi, D. G., C. Stathopoulos, et al. (2002). "Bacterial outer membrane ushers contain distinct targeting and assembly domains for pilus biogenesis." J Bacteriol **184**(22): 6260-6269.
- Tieng, V., C. Le Bouguenec, et al. (2002). "Binding of *Escherichia coli* adhesin AfaE to CD55 triggers cell-surface expression of the MHC class I-related molecule MICA." Proc Natl Acad Sci U S A **99**(5): 2977-2982.
- Tilley, R. J. D. (2006). Crystals and crystal structures. Hoboken, NJ, John Wiley.
- Todd, E. (1990). "Epidemiology of foodborne illness: North America." Lancet **336**(8718): 788-790.
- Touchon, M., C. Hoede, et al. (2009). "Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths." PLoS Genet **5**(1): e1000344.
- Tran Van Nhieu, G., R. Bourdet-Sicard, et al. (2000). "Bacterial signals and cell responses during *Shigella* entry into epithelial cells." Cell Microbiol **2**(3): 187-193.
- Varma, J. K., K. D. Greene, et al. (2003). "An outbreak of *Escherichia coli* O157 infection following exposure to a contaminated building." JAMA **290**(20): 2709-2712.
- Vetsch, M., D. Erilov, et al. (2006). "Mechanism of fibre assembly through the chaperone-usher pathway." EMBO Rep **7**(7): 734-738.
- Vetsch, M., C. Puorger, et al. (2004). "Pilus chaperones represent a new type of protein-folding catalyst." Nature **431**(7006): 329-333.
- Virji, M., K. Makepeace, et al. (1996). "Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic *neisseriae*." Mol Microbiol **22**(5): 941-950.
- Waksman, G. and S. J. Hultgren (2009). "Structural biology of the chaperone-usher pathway of pilus biogenesis." Nat Rev Microbiol **7**(11): 765-774.
- Wanger, G., T. C. Onstott, et al. (2008). "Stars of the terrestrial deep subsurface: a novel 'star-shaped' bacterial morphotype from a South African platinum mine." Geobiology **6**(3): 325-330.
- Wanke, C. A., J. Gerrior, et al. (1998). "Successful treatment of diarrheal disease associated with enterotoxigenic *Escherichia coli* in adults infected with human immunodeficiency virus." J Infect Dis **178**(5): 1369-1372.

- Wanke, C. A., H. Mayer, et al. (1998). "Enteraggative *Escherichia coli* as a potential cause of diarrheal disease in adults infected with human immunodeficiency virus." J Infect Dis **178**(1): 185-190.
- Wanke, C. A., J. B. Schorling, et al. (1991). "Potential role of adherence traits of *Escherichia coli* in persistent diarrhea in an urban Brazilian slum." Pediatr Infect Dis J **10**(10): 746-751.
- Wei, J., M. B. Goldberg, et al. (2003). "Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T." Infect Immun **71**(5): 2775-2786.
- Wells, J. G., B. R. Davis, et al. (1983). "Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype." J Clin Microbiol **18**(3): 512-520.
- Williams, D. T., Y. Chaudhry, et al. (2004). "Interactions of decay-accelerating factor (DAF) with haemagglutinating human enteroviruses: utilizing variation in primate DAF to map virus binding sites." J Gen Virol **85**(Pt 3): 731-738.
- Williams, P., Y. Chaudhry, et al. (2003). "Mapping CD55 function. The structure of two pathogen-binding domains at 1.7 Å." J Biol Chem **278**(12): 10691-10696.
- Winn, M. D., M. N. Isupov, et al. (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." Acta Crystallogr D Biol Crystallogr **57**(Pt 1): 122-133.
- Wolf, M. K. (1997). "Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*." Clin Microbiol Rev **10**(4): 569-584.
- Young, K. D. (2006). "The selective value of bacterial shape." Microbiol Mol Biol Rev **70**(3): 660-703.
- Zavialov, A. V., J. Berglund, et al. (2003). "Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation." Cell **113**(5): 587-596.