



# *University of* **HUDDERSFIELD**

## **University of Huddersfield Repository**

Dyer, Adam

Identification and Structural Characterisation of Novel Outer Membrane Proteins in the genus *Borrelia*

### **Original Citation**

Dyer, Adam (2013) Identification and Structural Characterisation of Novel Outer Membrane Proteins in the genus *Borrelia*. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/19265/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

**Identification and Structural Characterisation of Novel  
Outer Membrane Proteins in the genus *Borrelia***

Adam Dyer MSc

A thesis submitted to the University of Huddersfield  
in partial fulfilment of the requirements for  
the degree of Doctor of Philosophy

June 2013

## **Copyright declaration**

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

*For my family*



*“Our virtues and our failings are inseparable,  
Like force and matter.  
When they are separate,  
Man is no more.”*

Nikola Tesla, 1856 - 1943.

## Abstract

Although considered Gram-negative the outer membrane (OM) of *Borrelia* is unique, consisting of a variety of glycolipids, a significant number of surface exposed lipoproteins and a smaller number of integral membrane-spanning  $\beta$ -barrels. Many of these proteins are known to act as adhesins and invasins binding to regulators of the host immune response and extracellular matrix proteins. Many Gram-negative bacteria have the evolutionarily conserved OmpA-like transmembrane domain consisting of an eight-stranded membrane spanning  $\beta$ -barrel. The prototypical example is *E. coli* OmpA; a multifunctional protein involved in a wide variety of physiological and pathological functions. To date the OmpA-like transmembrane domain has not been identified in the Spirochaete phylum. An approach based on hidden Markov models and topology and fold prediction identified a group of homologous genes in the genus *Borrelia* (BAPKO\_0026, BAPKO\_0422, BAPKO\_0423 and BAPKO\_0571), which are predicted to form an eight stranded OM-spanning  $\beta$ -barrel structure with similar structure to *E. coli* OmpA/W. One of these orthologues in *B. afzelii* (BAPKO\_0422) has been expressed, purified and characterised *in vitro*; circular dichroism studies show a large percentage of  $\beta$ -strand. A low resolution molecular envelope generated from small-angle X-ray scattering data is consistent with an eight-stranded  $\beta$ -barrel.

## Table of Contents

Copyright declaration .....	2
Abstract .....	5
Table of Contents .....	6
List of Figures .....	10
List of Tables.....	13
List of Equations .....	14
Abbreviations .....	15
Acknowledgements .....	17
Chapter 1. Introduction.....	18
1.1 Lyme disease .....	18
1.1.1 History.....	18
1.1.2 Symptoms .....	19
1.1.3 Diagnosis, treatment & vaccines.....	25
1.1.4 Relapsing fever .....	27
1.2 The genus <i>Borrelia</i> .....	27
1.2.1 The genetics of <i>Borrelia</i> .....	30
1.2.2 The ultrastructure of <i>Borrelia</i> .....	32
1.2.3 The outer membrane of <i>Borrelia</i> .....	35
1.2.4 Immune evasion and infectivity of <i>Borrelia</i> .....	37
1.2.5 <i>Borrelia</i> implicated with Alzheimer's disease.....	48
1.3 <i>Ixodid</i> ticks .....	50
1.3.1 Tick life cycle.....	51
1.3.2 The influence of tick saliva on survival of <i>Borrelia</i> in hosts .....	53
1.4 The Complement System.....	58
1.4.1 Activation of the complement system.....	58
1.4.2 Evasion of the complement system by <i>Borrelia</i> and other bacteria.....	61
1.5 Outer Membrane Proteins.....	64
1.5.1 OmpA & its role in infection & immune evasion .....	65
1.5.2 OmpW .....	67
1.5.3 OmpX.....	68

1.5.4	A structural comparison of OmpA, OmpX and OmpW .....	69
1.6	Aims of the research .....	71
Chapter 2.	Materials & Methods .....	73
2.1	Materials .....	73
2.1.1	Reagents, buffers and purchased media .....	74
2.1.2	Primers .....	76
2.1.3	Plasmid vector.....	78
2.2	Methods .....	80
2.2.1	Nested PCR amplification and cloning of targets .....	80
2.2.2	Restriction digest and production of DNA constructs .....	80
2.2.3	Agarose gels.....	82
2.2.4	Competent cell preparation .....	82
2.2.5	Transformations .....	82
2.2.6	Extraction of plasmid constructs.....	83
2.2.7	Expression of protein in inclusion bodies .....	84
2.2.8	SDS-PAGE and Western Blot analysis.....	84
2.2.9	Purification of inclusion bodies from autoinduction cultures.....	86
2.2.10	Immobilised Metal Affinity Chromatography .....	87
2.2.11	Size Exclusion Chromatography.....	88
2.2.12	SEC-MALLS .....	90
2.2.13	Circular Dichroism.....	91
2.2.14	Small Angle X-Ray Scattering.....	92
2.2.15	X-Ray crystallography .....	95
Chapter 3.	Identification and bioinformatic studies of potential OmpA-like proteins in <i>Borrelia</i> . .....	99
3.1	Classification of known Omps .....	99
3.2	Profile hidden Markov model searches .....	105
3.3	Topology predictions .....	111
3.3.1	FFAS03 assessment .....	112
3.3.2	Signal peptide analysis.....	113
3.3.3	-barrel prediction.....	116
3.3.4	Summary .....	117
3.4	Homology modelling of OmpA-like proteins .....	121

3.5	Discussion.....	123
Chapter 4.	Cloning and preparation of OmpA-like protein for structural studies .....	130
4.1	Amplification of genes for cloning.....	130
4.2	Production of gene constructs.....	136
4.3	Recombinant Expression of BAPKO_0422 .....	139
4.3.1	Codon usage.....	139
4.3.2	Autoinduction.....	141
4.4	Purification and refolding of BAPKO_0422 .....	142
4.4.1	Purification of inclusion bodies .....	142
4.4.2	IMAC purification & refolding.....	144
4.4.3	SEC Purification .....	148
4.5	Discussion.....	150
Chapter 5.	Structural and functional studies of BAPKO_0422.....	153
5.1	Circular dichroism .....	153
5.1.1	Circular dichroism of BAPKO_0422.....	157
5.2	SEC-MALLS .....	165
5.3	Small Angle X-Ray Scattering .....	167
5.3.1	Control analysis (lysozyme).....	168
5.3.2	Experimental analysis, BAPKO_0422.....	178
5.4	Crystallisation trials of BAPKO_0422 .....	192
5.4.1	Known crystallisation of TM -barrel proteins from <i>E. coli</i> .....	193
5.4.2	Vapour diffusion crystallisation trials.....	194
5.4.3	<i>In meso</i> crystallography .....	198
5.5	Discussion.....	202
Chapter 6.	Conclusions & Future Directions.....	210
6.1	Conclusions .....	210
6.2	Future directions .....	213
Chapter 7.	References.....	215
Chapter 8.	Appendix.....	229
8.1	Redundancy reduced lists of known OM proteins .....	229

8.2	Sequence alignments used for homology model generation .....	236
8.3	Conditions for membrane protein crystallisation screen .....	237

## List of Figures

Figure 1.1. Erythema Migrans. ....	19
Figure 1.2. Lyme disease in England. ....	24
Figure 1.3. 16s rRNA Phylogenetic tree of Spirochaetes. ....	29
Figure 1.4. Geographical distribution of various <i>Borrelia</i> species. ....	30
Figure 1.5. Basic diagram of the ultrastructure of a <i>Borrelia</i> Spirochaete ....	33
Figure 1.6. Basic diagram showing the anchoring points of endoflagella. ....	33
Figure 1.7. Anchoring of endoflagella in the cytoplasmic membrane of <i>Borrelia</i> . ....	34
Figure 1.8. Cross section of a <i>Borrelia bacterium</i> . ....	34
Figure 1.9. Schematic representation of known lipoproteins and outer membrane proteins in <i>Borrelial</i> membranes. ....	35
Figure 1.10. Glycolipids present in the OM of <i>Borrelia</i> . ....	36
Figure 1.11. Cholesterol rich lipid rafts in the outer membrane of <i>Borrelia</i> . ....	37
Figure 1.12. Crystal structure of BbCRASP-1. ....	39
Figure 1.13. Crystal structures of OspA from <i>B. burgdorferi</i> s.s. ....	41
Figure 1.14. Crystal structure of the 16.5 kDa C terminus of OspB from <i>B. burgdorferi</i> s.s..	43
Figure 1.15. Comparison of the C-termini of OspA and OspB. ....	43
Figure 1.16. Crystal structure of OspC from <i>B. burgdorferi</i> s.s. ....	45
Figure 1.17. <i>Ixodes ricinus</i> suitable habitats in the UK and Europe. ....	50
Figure 1.18. A female and male <i>I. ricinus</i> . ....	51
Figure 1.19. The two year life cycle of an <i>Ixodes</i> tick. ....	52
Figure 1.20. The anterior morphology of <i>I. ricinus</i> . ....	53
Figure 1.21. <i>Ixodid</i> tick hypostome. ....	54
Figure 1.22. Tick protein interactions with <i>Borrelia</i> and mammalian host evasion. ....	57
Figure 1.23. Activation of the complement immune system by three different mechanisms. ....	60
Figure 1.24. Evasion of complement response by pathogens. ....	61
Figure 1.25. Representation of the domain structure of factor H. ....	63
Figure 1.26. The TM spanning domain of <i>E. coli</i> OmpA. ....	65
Figure 1.27. <i>E. coli</i> OmpW. ....	68
Figure 1.28. <i>E. coli</i> OmpX. ....	69
Figure 1.29. OmpA, W and X from <i>E. coli</i> . ....	70
Figure 1.30. The aromatic girdle of OmpA, OmpW and OmpX. ....	71
Figure 2.1. Restriction sites present throughout the gene BAPKO_0422. ....	78

Figure 2.2. Plasmid map of pET-47b(+) (Novagen).....	79
Figure 2.3. The cloning and expression region of pET-47b(+) .....	79
Figure 2.4. Western blot module set up. ....	86
Figure 2.5. SEC calibration curve.....	89
Figure 2.6. Vapour diffusion setups.....	95
Figure 2.7. Detergent binding by membrane proteins. ....	96
Figure 3.1. An example of a redundancy reduced protein search.....	100
Figure 3.2. NewHampshire Tree produced using a list of known TM $\alpha$ -barrel proteins reduced by 40 %.....	101
Figure 3.3. NewHampshire Tree produced using a list of known TM $\alpha$ -barrel proteins reduced by 90 %.....	102
Figure 3.4. Sequence alignment of <40 % identical known TM $\alpha$ -barrel proteins.....	104
Figure 3.5. A small HMM profile.....	106
Figure 3.6. HMM logo of known TM $\beta$ -barrel protein sequences with a 40 % redundancy reduction. ....	107
Figure 3.7. Signal peptidases found in Gram-negative bacteria. ....	114
Figure 3.8. SignalP 4.1 prediction of OmpA-like protein signal sequences.....	115
Figure 3.9. TM $\alpha$ -barrel predictions of OmpA-like proteins. ....	116
Figure 3.10. Summary of topology prediction.....	118
Figure 3.11. HMM logo of putative <i>Borrelial</i> OmpA-like proteins. ....	120
Figure 3.12. Sequence alignments of BAPKO_0422 with known <i>E. coli</i> OM-spanning proteins.....	121
Figure 3.13. BAPKO_0422 Models produced using <i>E. coli</i> OM-spanning proteins as templates. ....	122
Figure 4.1. The amplification of a target gene by nested PCR. ....	131
Figure 4.2. Amplification of PepX, UvrA and BG0027 using nPCR.....	132
Figure 4.3. Gel of 35 cycle PCR using secondary primers only.....	133
Figure 4.4. Amplification screen for BG0407 and BAPKO_0422.....	134
Figure 4.5. An example the positioning of a forward overhang primer, A0422CF.....	135
Figure 4.6. Results of BAPKO_0422 sequencing. ....	135
Figure 4.7. Amplified BAPKO_0422 BLAST result. ....	136
Figure 4.8. Digested gene constructs. ....	137
Figure 4.9. BAPKO_0422 gene sequence obtained from construct. ....	138
Figure 4.10. Translation of sequenced constructs.....	138



Figure 4.11. Top hit from the BLAST search of BAPKO_0422. ....	139
Figure 4.12. <i>E. coli</i> K-12 codon usage in the <i>B. afzelii</i> gene BAPKO_0422. ....	140
Figure 4.13. Expression of BAPKO_0422. ....	142
Figure 4.14. Separation of inclusion bodies.....	143
Figure 4.15. IMAC Chromatogram of purification of BAPKO_0422 in inclusion bodies solubilised with 8 M urea.....	144
Figure 4.16. IMAC Chromatogram of purification of BAPKO_0422 in inclusion bodies solubilised with 6 M GuHCl. ....	145
Figure 4.17. IMAC Chromatogram of purification of BAPKO_0422 in inclusion bodies solubilised with 6 M GuHCl with the addition of 10 % (v/v) glycerol. ....	146
Figure 4.18. IMAC Purification and on-column refold of BAPKO_0422. ....	148
Figure 4.19. SEC Chromatogram of BAPKO_0422.....	149
Figure 4.20. SDS-PAGE of purified BAPKO_0422. ....	149
Figure 5.1. The CD effect. ....	154
Figure 5.2. Addition of L and R polarised light of different magnitudes. ....	155
Figure 5.3. Example CD spectra. ....	156
Figure 5.4. CD spectra of <i>E. coli</i> OmpA TM domain.....	157
Figure 5.5. CD of 0.33 mg/ml BAPKO_0422. ....	158
Figure 5.6. Plots of BAPKO_0422 DichroWeb analysis.....	159
Figure 5.7. CD spectra of 0.166 mg/ml BAPKO_0422. ....	160
Figure 5.8. CDSSTR analysis of BAPKO_0422 at 0.166 mg/ml.....	161
Figure 5.9. CD spectra of BAPKO_0422 at various temperatures. ....	162
Figure 5.10. Predicted CD spectra of known OM proteins from <i>E. coli</i> . ....	164
Figure 5.11. SEC-MALLS trace of BAPKO_0422. ....	165
Figure 5.12. The sine wave of an X-ray.....	167
Figure 5.13. Scattering angle relative to particle size. ....	168
Figure 5.14. Scattering produced by 10 mg/ml Lysozyme in 0.1 and 0.01 % (w/v) LDAO. ....	169
Figure 5.15. GNOM reciprocal space fit of 10 mg/ml lysozyme. ....	172
Figure 5.16. $P(r)$ of lysozyme in 0.1 % (w/v) LDAO with a $D_{\max}$ of 50.....	173
Figure 5.17. $P(r)$ plots of lysozyme in 0.1 % (w/v) LDAO with $D_{\max}$ values 40 & 60.....	173
Figure 5.18. Molecular envelopes of lysozyme generated from 10 mg/ml lysozyme in SAXS 0.1 % (w/v) LDAO buffer.....	174
Figure 5.19. Lysozyme molecular envelope after refinement. ....	176
Figure 5.20. DAMMIN refinement of the molecular envelope. ....	177

Figure 5.21. Scattering pattern produced by BAPKO_0422 in solution. ....	178
Figure 5.22. SAXS data obtained from 6 mg/ml BAPKO_0422.....	179
Figure 5.23. SAXS data obtained from 10 mg/ml BAPKO_0422.....	180
Figure 5.24. GNOM fit and $P(r)$ plots of BAPKO_0422 datasets.....	183
Figure 5.25. Molecular envelopes generated from dataset 6TT. ....	184
Figure 5.26. 6TT molecular envelopes after refinement.....	186
Figure 5.27. DAMMIN refined molecular envelope 6TT. ....	187
Figure 5.28. Molecular envelopes generated from dataset 10NT.....	188
Figure 5.29. 10NT data averaged molecular envelopes.....	190
Figure 5.30. DAMMIN refined molecular envelope 10NT.....	191
Figure 5.31. Theoretical protein crystallisation phase diagram.....	192
Figure 5.32. Salt crystals obtained from CSS I and II. ....	196
Figure 5.33. Crystals grown by vapour diffusion. ....	197
Figure 5.34. Formation of crystals in the lipid cubic phase.....	199
Figure 5.35. Crystals produced using <i>In meso</i> crystallisation. ....	200
Figure 5.36. Crystals grown from <i>In meso</i> crystallography. All drops were set up using precipitant A9.....	201

## List of Tables

Table 1.1. Various presentations of Lyme disease at each stage of infection. ....	23
Table 2.1. Chemicals and reagents used in this research. ....	74
Table 2.2. Buffers and Media used throughout this research. ....	75
Table 2.3. PCR reagents, enzymes and buffers used in this work. ....	76
Table 2.4. Primers for the nPCR amplification of housekeeping genes PepX and UvrA. ....	76
Table 2.5. Primers for the amplification of target genes.....	77
Table 2.6. Primers for the addition of restriction sites.....	78
Table 2.7. Detergents and lipid commonly used in membrane protein purification and crystallisation. ....	97
Table 3.1. BLAST search of 1BXW from <i>E. coli</i> against the <i>B. garinii</i> proteome. ....	99
Table 3.2. HMM profile search results of <i>Borrelial</i> proteomes. ....	110
Table 3.3. Summary of potential OmpA-like proteins identified in <i>Borrelia</i> . ....	111
Table 3.4. FFAS03 results from BG0407. ....	112
Table 3.5. Summary of FFAS03 results.....	113

Table 3.6. Results of SignalP 4.0 and PRED-TMBB analysis. ....	117
Table 3.7. Putative small TM $\alpha$ -barrel proteins in <i>Borrelia</i> . ....	117
Table 5.1. Results from DichroWeb analysis. ....	161
Table 5.2. DichroWeb analysis of BAPKO_0422 at varying temperatures. ....	163
Table 5.3. Calculated R <sub>g</sub> of lysozyme. ....	170
Table 5.4. DAMSEL comparison of lysozyme molecular envelope solutions. ....	175
Table 5.5. PRIMUS and GNOM analysis. ....	182
Table 5.6. DAMSEL comparison of 6TT molecular envelope solutions. ....	185
Table 5.7. DAMSEL comparison of 10NT molecular envelope solutions. ....	189
Table 5.8. Known crystallisation conditions of $\alpha$ -barrel proteins. ....	193
Table 5.9. CSS I screen. ....	194
Table 5.10. CSS II screen. ....	195
Table 5.11. Results of initial crystal screening. ....	195

## List of Equations

Equation 1. Calculating the concentration of DNA. ....	81
Equation 2. Calculating DNA purity. ....	81
Equation 3. Calculation of $A_{260}/A_{280}$ . ....	155
Equation 4. Small angle scattering as a function of $q$ . ....	168
Equation 5. Calculating the radius of gyration for a spherical object from a Guinier plot. ....	170

## Abbreviations

AP	Alternative Pathway
BBB	Blood-brain barrier
BbCRASP	<i>B. burgdorferi</i> Complement Regulator Acquiring Surface Protein
BLAST	Basic Local Alignment Search Tool
BMEC	Brain micro vascular endothelial cells
C <sub>8</sub> E <sub>4</sub>	n-octyltetraoxyethylene
CD	Circular Dichroism
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CM	Cytoplasmic membrane
CNS	Central nervous system
CP	Classical Pathway
Dbp	Decorin binding protein
DDM	n-Dodecyl -D-maltoside
Erp	OspE Related Protein
fH	Factor H
fHR-1	Factor H-related protein
Fn	Fibronectin
FPLC	Fast Protein Liquid Chromatography
HMM	Hidden-Markov Model
IMAC	Immobilised Metal Affinity Chromatography
LCP	Lipid Cubic Phase
LDAO	N,N-Dimethyldodecylamine <i>N</i> -oxide
LPS	Lipopolysaccharide
LS	Light Scattering
MBL	Mannose-binding lectin
MS	Mass Spectrometry
MSA	Multiple Sequence Alignment
MO	Monoolein
NMR	Nuclear Magnetic Resonance Spectroscopy
OG	n-octyl- -D-glucoside
OM	Outer membrane
Omp	Outer membrane protein

Osp	Outer surface protein
PEG	Polyethylene Glycol
Rg	Radius of Gyration
RI	Refractive Index
SAXS	Small-Angle X-ray Scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
SEC-MALLS	Size Exclusion Chromatography-Multi Angle Laser Light Scattering
s.s	<i>Sensu stricto</i>
s.l.	<i>Sensu lato</i>

## Acknowledgements

I would like to thank Dr. Richard Bingham for sharing his expertise and skills, for his support, encouragement and advice both scientific and life related.

I would also like to thank Dan and James for being there to bounce ideas off, for the helpful suggestions and for the countless laughs and good times; may there be many more to come!

Thanks to Dr Peter Laity (University of Huddersfield) for assistance with collection and interpretation of SAXS data, to Dr Andrew Leech (University of York) for assistance with Circular Dichroism, to Dr Arwen Pearson (University of Leeds) for assistance with X-Ray diffraction, to Glenn Robinson for assistance preparing the lipid cubic phase for crystallography and to the biological science technicians Felix Owusu-Kwarteng, Maggie Craven, Sophie Sinclair and Elena Tillotson for all their help and patience over the years. Thanks to the University of Huddersfield for the fee waiver scholarship.

Thanks to Dr Gabi Margos (University of Bath) and Dr Volke Fingerle & Cecilia Hizo-Teufel for *Borrelial* DNA samples.

Finally a huge heartfelt thank you to my parents, Alan and Gwyneth for their unwavering love, support and encouragement throughout the years, without which, none of this would have been possible.

## Chapter 1. Introduction

### 1.1 Lyme disease

#### 1.1.1 History

Lyme disease, also known as Lyme borreliosis (Franz and Krause, 2003) is a tick-borne zoonosis that affects multiple systems in mammalian hosts; the etiological agent being three species of the *Borrelia* genus: *B. afzelii*, *B. burgdorferi* s.s. and *B. garinii* (Burgdorfer *et al.*, 1982, Franz and Krause, 2003, Kurtenbach *et al.*, 2006, Rupprecht *et al.*, 2008). The initial symptom, an erythema migrans rash was first described in Europe in 1910 and was identified to be caused by an organism transmitted by ticks, while the common later stage symptoms of Lyme disease were first characterised in Europe in 1883. Later in 1922 symptoms related to the CNS were recorded, again in Europe (Athreya and Rose, 1996) (Franz and Krause, 2003). The recent history of Lyme disease began with an outbreak in the early 1970's in the USA (Steere *et al.*, 1977, Steere *et al.*, 2004). The transmission vector was identified as species of the *Ixodes* tick around the same time (Steere *et al.*, 1978, Alitalo, 2004) although it was not until 1982 however that the responsible Spirochaete *B. burgdorferi* s.s. was first isolated, identified and confirmed as the cause of Lyme disease (Burgdorfer *et al.*, 1982, Hayes *et al.*, 1983, Johnson *et al.*, 1984, Burgdorfer *et al.*, 1986).

Lyme disease is the most prevalent tick-borne disease in humans in the Northern Hemisphere with 100-130 reported cases per 100,000 people in the UK and 20-100 reported cases per 100,000 people in the US (Rupprecht *et al.*, 2008). The disease was named after the town of Old Lyme, Connecticut, U.S.A. where seemingly epidemic outbreaks of juvenile rheumatoid arthritis had been occurring annually in the summer and autumn since at least 1972 (Steere *et al.*, 1977). The results of a thorough clinical investigation into the outbreak revealed that a number of patients had a bulls-eye-like rash known as an erythema migrans (EM) and had been bitten by a tick prior to the onset of arthritis (Steere *et al.*, 1978). As the disease progressed, more severe symptoms presented such as the aforementioned arthritis, skin conditions, and neurological and cardiac symptoms (Steere *et al.*, 1977, Steere *et al.*, 1978, Burgdorfer *et al.*, 1982, Franz and Krause, 2003). It was from this and several other studies that the agent responsible for the outbreak was identified as a Spirochaete of the genus *Borrelia* and named after Dr Willy Burgdorfer, one of the discoverers (Steere *et al.*, 1977, Steere *et al.*, 1978, Burgdorfer *et al.*, 1982, Johnson *et al.*, 1984). Species of *Borrelia* have

also been found in the midgut of mosquitoes, *Culex Pipiens* and *Aedes vexans* (Krupka *et al.*, 2007).

### 1.1.2 Symptoms

Lyme disease caused by mammalian infection with species of *Borrelia* produces a wide variety of symptoms that differ depending on the species and strain of bacteria causing the infection (Rosa *et al.*, 2005, Hildenbrand *et al.*, 2009). The symptoms exacerbate over time affecting multiple tissues and systems in the host's body, notably joints, the CNS and myocardium (Comstock *et al.*, 1993, Kurtenbach *et al.*, 2006, Rupprecht *et al.*, 2008). Lyme disease has three stages of infection: initial, acute and persistent (or chronic) (Rosa *et al.*, 2005).

During the initial infection phase of *Borrelial* infection of a human host the first symptom to be presented is an erythema migrans (EM), a characteristic round/elliptical bulls-eye like rash that occurs at the site of first infection/the site of the tick bite (Figure 1.1) (Rosa *et al.*, 2005). This occurs in 60 to 80 % of confirmed infected patients (Alitalo, 2004, Auwaerter, 2007, Krupka *et al.*, 2007, Rupprecht *et al.*, 2008, Fallon *et al.*, 2010) and can last up to twelve weeks (Franz and Krause, 2003). This initial localised symptom is common to the three major Lyme disease causing species of *Borrelia* with the exception that an EM from *B. burgdorferi* s.s. has been known to clear in the middle, leaving a ring shaped rash (Franz and Krause, 2003). Other initial symptoms are flu-like in nature, ranging from fever and chills to headaches, lasting up to several weeks (Hyde and Johnson, 1984, Athreya and Rose, 1996, Hildenbrand *et al.*, 2009), although these are most common in the USA (Steere *et al.*, 2004). Half of the patients infected with *B. burgdorferi* s.s. develop multiple secondary EMs similar to the initial rash but smaller (Franz and Krause, 2003).



**Figure 1.1. Erythema Migrans.** An image of an erythema migrans with a clear region in the centre. File licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.



If left untreated, symptoms of the second (acute) stage of infection begin to manifest several weeks to months after first infection once the *Spirochaete* has disseminated to tissues throughout the body (Steere *et al.*, 2004, Hildenbrand *et al.*, 2009). Broadly, symptoms include inflammation in: joints (arthritis), muscles including the heart (carditis), both the central and peripheral nervous systems, and Bell's palsy, musculoskeletal pain and acute lymphocytic meningitis (Athreya and Rose, 1996, Steere *et al.*, 2004, Fallon *et al.*, 2010). Unlike the initial stage of infection, the symptoms presented during the second stage of infection show greater differentiation between the *Borrelial* species (Franz and Krause, 2003), with arthritic symptoms most common in the USA (*B. burgdorferi* s.s.) and neurological symptoms more prevalent in Europe (*B. afzelii* & *B. garinii*) (Rosa *et al.*, 2005).

Lyme arthritis is the most common symptom of the second and third stages of infection in the USA with approximately 60 % of patients developing symptoms (Steere *et al.*, 2004) compared to a much smaller proportion of patients in Europe (Franz and Krause, 2003). In contrast, Lyme carditis has only been reported in a minority of cases of Lyme disease, typical symptoms include atrial fibrillations and tachycardia (Franz and Krause, 2003).

Neurological symptoms (known as neuroborreliosis) that occur in the second stage of Lyme disease are most commonly caused by the European strains of *Borrelia*: *B. afzelii* and *B. garinii* (Franz and Krause, 2003). Symptoms include Bannwarth's syndrome (radiculo neuropathy with meningitis), cranial neuropathy, Bell's palsy, encephalitis and inflammation in the CNS (Kalish *et al.*, 1993, Franz and Krause, 2003, Livengood and Jr, 2006, Fallon *et al.*, 2010).

Stage three of infection (chronic or persistent infection) presents the most severe symptoms, often second stage symptoms become chronic and much more serious. Chronic Lyme arthritis can occur in joints, lasting up to a year or longer. Synovial lesions that are similar to other forms of chronic inflammatory arthritis including rheumatoid arthritis have been found to be present in patients (Kalish *et al.*, 1993).

The skin condition acrodermatitis chronic atrophans has been found in patients years after they were infected with *B. afzelii* (Steere *et al.*, 2004, Krupka *et al.*, 2007). Neuroborreliosis can progress to a chronic illness, the symptoms of which include dementia, mental disorders, cranial neuropathy, synovial lesions, cognitive abnormalities and chronic encephalitis and

encephalomyelitis, leading to hemiparesis (Kalish *et al.*, 1993, Franz and Krause, 2003, Fallon *et al.*, 2010). Lyme encephalopathy, a syndrome that causes spinal radicular pain and cognitive disturbances has been described to occur in rare instances (Steere *et al.*, 2004). It has been shown that a small number of patients may develop post-Lyme disease syndrome (or chronic fatigue borreliosis); an ailment similar in nature to chronic fatigue syndrome or fibromyalgia and includes symptoms such as severe musculoskeletal pain and fatigue (Steere *et al.*, 2004). A summary of symptoms is included in Table 1.1.

System	Early infection		Acute Infection		Chronic infection		Species	References
	Stage 1 (Localised)		Stage 2 (Disseminated)		Stage 3 (Persistent)			
Skin	Localised Erythema migrans (EM)		Malar rash, lymphocytoma, secondary EM, multiple EM, annular skin lesions		Rarely encountered		<i>Bb</i>	(Rupprecht et al., 2008, Auwaerter, 2007, Alitalo, 2004, Fallon et al., 2010, Franz and Krause, 2003, Steere et al., 2004, Krupa et al., 2007)
			Borrelial lymphocytoma		Acrodermatitis chronic atrophans, Scleroderma-like lesions		<i>Ba</i> <i>Bg</i>	(Franz and Krause, 2003, Steere et al., 2004, Krupa et al., 2007)
Musculoskeletal	-		Migratory pain in joints, tendons, muscle, bone, brief spells of arthritis, inflammation in muscles, gait disturbance and muscle weakness		Chronic Lyme arthritis, joint inflammation, treatment resistant arthritis, chronic encephalomyelitis (chronic fatigue syndrome), spastic paraplegia		<i>Bb</i>	(Fallon et al., 2010, Kalish et al., 1993, Franz and Krause, 2003, Rosa et al., 2005, Krupa et al., 2007, Kaiser et al., 1997)
			Infrequent arthritis, inflammation in joints, not as severe as <i>Bb</i> arthritis.		Rarer cases of persistent arthritis		<i>Ba</i> <i>Bg</i>	(Franz and Krause, 2003)
Cardiovascular	-		Mild myocarditis		Mild myocarditis		<i>Bb</i>	(Fallon et al., 2010, Franz and Krause, 2003)
			Myocarditis, pericarditis, inflammation in heart		Dilated cardiomyopathy		<i>Ba</i> <i>Bg</i>	(Franz and Krause, 2003, Rosa et al., 2005)
Nervous & Neurological	-		Meningitis		Slight cognitive disruptions, not as severe as <i>Ba</i> and <i>Bg</i> symptoms		<i>Bb</i>	(Franz and Krause, 2003)
			Aseptic meningitis - severe		Chronic encephalitis, Metamorphopsia		<i>Ba</i>	(Fallon et al., 2010, Kalish

	in children, Bell's palsy, encephalitis, inflammation in CNS, acute lymphocytic meningitis	dementia, mental disorders, cranial neuropathy, synovial lesions, cognitive abnormalities, Bannwarth's syndrome, cranial neuropathy	<i>Bg</i>	et al., 1993, Franz and Krause, 2003, Steere et al., 2004, Pachner, 2006, Halperin, 2003)
<b>General symptoms</b>	Flu-like symptoms – fever, chills, headache	Chronic fatigue borreliosis, Fatigue	All	(Hyde and Johnson, 1984, Hildenbrand et al., 2009, Steere et al., 2004, Rosa et al., 2005)
<b>Unusual symptoms</b>	Neuroborreliosis presenting as Alice in Wonderland syndrome	Auditory hallucinations Metamorphopsia	Not specified	(Binalsheikh et al., 2012)
	Dysphagia, bad breath, fever	Tonsillitis	<i>Bv</i>	(Biswas and Stafford, 2010)
	Motor neurone disease	Chronic encephalomyelitis	<i>Bb</i>	(Kaiser et al., 1997)

**Table 1.1. Various presentations of Lyme disease at each stage of infection.** Initial data acquired from (Athreya and Rose, 1996) with additional data compiled from a variety of sources as detailed in the table. The species of *Borrelia* primarily responsible for the symptoms has been included: *Ba* = *B. afzelii*, *Bb* = *B. burgdorferi* s.s., *Bg* = *B. garinii*, *Bv* = *B. vincenti*.

Figure 1.2 shows confirmed cases of Lyme disease in England and Wales over a twelve year spell in the 1980's and 1990's. The majority of confirmed cases is in a forested area in the South, this is due to the *Ixodes* tick's preferred habitat being forested areas (section 1.3), the slightly warmer climate in the South and the abundance of wildlife in forests providing a natural host reservoir.



**Figure 1.2. Lyme disease in England.** Highlighted on the map is the Hampshire/New Forest area, the region with the most reported cases of indigenously acquired Lyme disease in England and Wales from 1986 to 1998. Data acquired from (Smith *et al.*, 2000).

### 1.1.3 Diagnosis, treatment & vaccines

Diagnosis of Lyme disease is a challenging proposition due to (with the exception of EM) the similarity of the initial symptoms to other diseases; in particular chronic fatigue syndrome, of which Lyme disease can be a cause (Kaiser *et al.*, 1997). Lyme disease is sometimes indistinguishable from other diseases such as Bannwarth's syndrome, hence Lyme disease has been referred to as the new 'great imitator' (Pachner, 2006). The diagnosis is made more clear by the presence of a visible EM and accompanying symptoms, however the diagnosis can be highly subjective (Aronowitz, 2012). In order to diagnose Lyme disease in the absence of a visible EM, the patient needs to produce a positive antibody response on a whole-cell *B. burgdorferi* lysate ELISA and Western Blot, be recognised as having symptoms characteristic of Lyme disease (Steere *et al.*, 2004) and show an increased white blood cell count (~ 100 white blood cells per cubic millimetre) in the cerebrospinal fluid (pleocytosis) (Alexander and Cox, 1995, Mygland *et al.*, 2010). An enzyme-linked immunosorbent assay has been developed to recognise several *Borrelial* proteins including: DbpA, BBK32 and OspC (sections 1.2.4.4 & 1.2.4.5), providing an overall sensitivity of 80 % and a specificity of 100 % (Skogman *et al.*, 2008).

If the infection is diagnosed in the early stage of infection or dissemination (less than 6 months), the recommended EFNS treatment is an oral course of doxycycline or an intravenous (IV) course of ceftriaxone in adults and amoxicillin in children (Steere *et al.*, 2004, Mygland *et al.*, 2010) over a period of two to four weeks (Rosa *et al.*, 2005), although if administered within 72 hours a single dose of doxycycline reduces the risk of Lyme disease by around 87 % (Nigrovic and Thompson, 2007). Chronic Lyme disease has been shown to require multiple, long term courses of antibiotics (Aronowitz, 2012), although even this may not be effective due to the Spirochaetes ability to invade the CNS (Livengood and Jr, 2006). Early CNS Lyme disease should be treated with a course of IV ceftriaxone (Mygland *et al.*, 2010). If symptoms persist for over six months, the disease is classified as post-Lyme disease syndrome; on which antibiotics have no effect (Mygland *et al.*, 2010).

Vaccines for Lyme disease have not successfully been produced due to the high variance in expression of antigenic *Borrelial* surface lipoproteins throughout infection (Krupka *et al.*, 2011). Two anti-OspA (see section 1.4.2.4) vaccines for *B. burgdorferi* s.s were developed in the early 1990's (Burkot *et al.*, 1994, Li *et al.*, 1997, Steere *et al.*, 1998) with the intention of

preventing Spirochaete transmission from arthropods to humans (Aronowitz, 2012). Both were found to be effective in animals and were deemed safe for humans and although one of these did not make it past phase III testing, the other was approved for general use (Steere *et al.*, 1998, Nigrovic and Thompson, 2007, Aronowitz, 2012). Mice immunised with recombinant *B. burgdorferi* s.s. OspA were protected against infection by the Spirochaete (Burkot *et al.*, 1994, Li *et al.*, 1997); additionally, monoclonal antibodies for the C-terminal end of the protein were also shown to protect the mouse from *Borrelial* invasion, which suggested that OspA binding was important for survival in mammalian invasion (Huang *et al.*, 1998) although it has been shown that this is not the case (Comstock *et al.*, 1993, Schwan and Piesman, 2000, Yang *et al.*, 2004, Pal *et al.*, 2004a). This vaccine did not make it to production.

The commercially available OspA vaccine was developed in 1998 (Steere *et al.*, 1998) and comprised full-length recombinant OspA with the lipid-chain added post-translation and an aluminium adjuvant added. The vaccine was administered in three 30 µg doses (initial, one month and twelve months); one month after the second dose 95 % of patients tested positive for an OspA antibody. One month after the third and final injection (13 months after the study began) 99 % of study participants tested positive for the OspA antibody. The study found that the three injections of OspA vaccine had an efficacy of approximately 76 % against *B. burgdorferi* s.s. (Steere *et al.*, 1998, Nigrovic and Thompson, 2007).

The vaccine had to be withdrawn three years later due to a decrease in sales following controversy over side effects (Krupka *et al.*, 2011); these included musculoskeletal pains and arthritis, thought to be due to a cross-reactive autoimmune response to OspA (Nigrovic and Thompson, 2007). A study in 2005 identified three cross-reactive neuronal human epitopes that shared identical amino-acid sequences with BbOspA; antibodies generated against these OspA homologous peptides reacted with human neurones in the brain spinal cord and dorsal root ganglia (Alaedini and Latov, 2005). As the perception of risks associated with the vaccine was high, the demand for the vaccine decreased (Nigrovic and Thompson, 2007). In addition, *Borrelia* downregulate OspA before entering the mammalian host, hence removing the antigenic surface that the antibodies stimulated by the vaccine would recognise (Cassatt *et al.*, 1998, Pal *et al.*, 2000, Schwan and Piesman, 2000, Yang *et al.*, 2004).

### 1.1.4 Relapsing fever

Relapsing fever (RF) is a tick and louse-borne zoonosis, caused by *B. hermsii*, *B. duttonii* and *B. recurrentis*. The two former species are spread by ticks of the genus *Ornithodoros*, while the latter is louse-borne and spread by *P. humanus* (Ras *et al.*, 1996), although there is phylogenetic evidence to suggest that *B. duttonii* and *B. recurrentis* are the same species (Cutler *et al.*, 1997). RF is characterised by recurring bouts of high bacteraemia, each caused by distinct populations of Spirochaete generated by the rearrangement of the genes for outer surface virulence proteins (Vuyyuru *et al.*, 2011). RF has been shown to recur at least 10 times without treatment (Bryceson *et al.*, 1970) and has an incubation period of 4 to 18 days; characteristic symptoms include headache, myalgia, chills, vomiting and abdominal pains (Schwan *et al.*, 2003). Secondary symptoms of RF include abdominal pains, tachycardia and tachypnoea (a high respiratory rate).

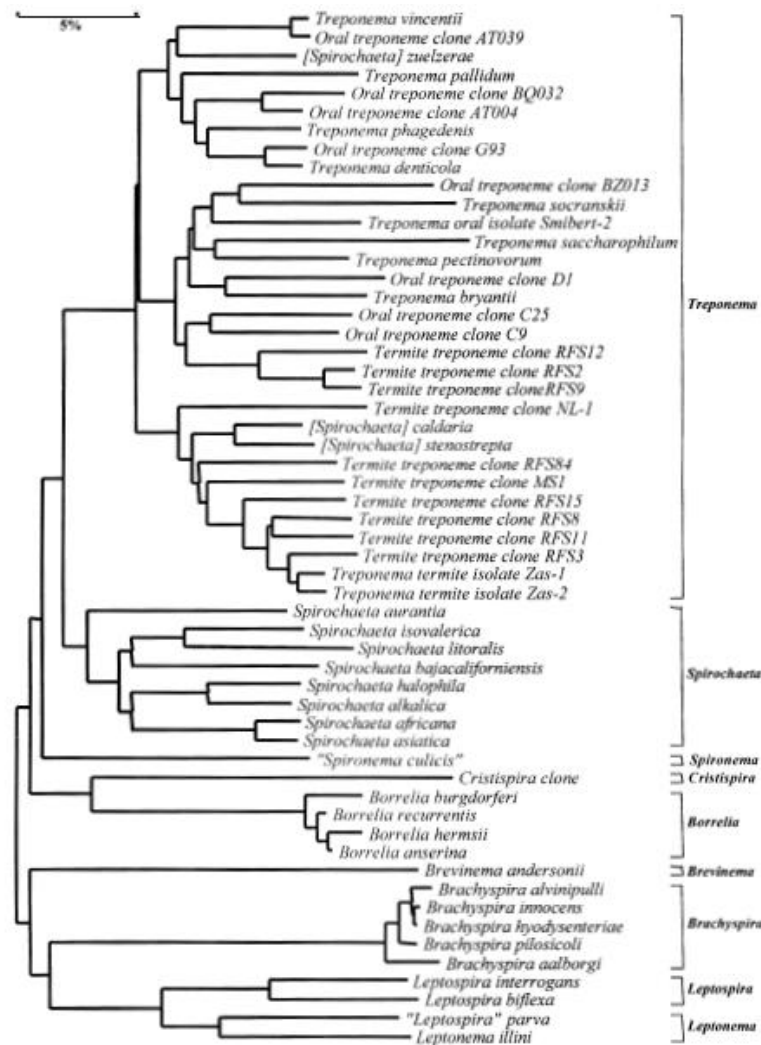
## 1.2 The genus *Borrelia*

The genus *Borrelia*, named after the French bacteriologist Amédée Borrel was first identified in 1834 by Ehrenberg under the name *Spirillum giganteum*. The name was changed to *Borrelia* in 1907, and was conferred by the Dutch bacteriologist Nicholas Swellengrebel (Wright, 2009). Species of *Borrelia* are the etiological agents of two major diseases in humans: the zoonosis Lyme disease (also known as Lyme borreliosis) and Relapsing Fever (also known as Relapsing fever borreliosis) (Cutler, 2009). *Borrelia burgdorferi* was named for the scientist Willy Burgdorfer, who, along with Allen Steere recognised that *Borrelia* was the causative agent of Lyme disease (Steere *et al.*, 1978, Burgdorfer *et al.*, 1982, Burgdorfer *et al.*, 1986, Burgdorfer, 1984). There are upwards of thirty known species of *Borrelia* detailed in various publications (Figure 1.3); and new species such as *B. americana* sp. nov., *B. kurtenbachii* sp. nov. and *B. carolinensis* sp. nov. have been identified in recent years (Rudenko *et al.*, 2010, Rudenko *et al.*, 2011, Margos *et al.*, 2011). *B. afzelii*, *B. burgdorferi* s.s and *B. garinii*, collectively known along with fifteen other species as *B. burgdorferi* s.l. (Rizzoli *et al.*, 2011), are the main species responsible for Lyme Disease in humans, while *B. duttonii*, *B. recurrentis* and *B. hermsii* are the main species known to cause Relapsing Fever (Schwan *et al.*, 2003, Bárcena-Uribarri *et al.*, 2010), although there is evidence that *B. recurrentis* is actually a strain of *B. duttonii* with a genome that appears to be decaying and losing proteins (Lescot *et al.*, 2008). *Borrelia* species responsible for causing Lyme Disease are transmitted to mammalian hosts during tick feeding, primarily by ticks of the *Ixodes*



genus: *I. ricinus*, *I. pacificus* and *I. dammini* (Burgdorfer *et al.*, 1982, Burgdorfer *et al.*, 1986). Relapsing fever Spirochaetes have been shown to be transmitted by hard ticks of the genera *Dermacentor*, *Amblyomma* and *Ornithodoros* (Athreya and Rose, 1996), the most common of which appear to be *B. hermsii* and *B. parkeri* transmitted by the tick species *O. hermsi* and *O. parkeri* respectively (Schwan *et al.*, 2003). Other species of *Borrelia* known to cause Relapsing Fever include: *B. turicatae*, *B. anserine*, *B. coriaceae*, *B. persica*, *B. hispanica* and *B. crocidurae* (Ras *et al.*, 1996). It is important to note however, not all strains of *Borrelia* are infectious in humans.

The phylum Spirochaetes consists of only one class and this class of one order. Differentiation of the Spirochaetes begins when the order is divided into three families - Spirochaetaceae, Brachyspiraceae and Leptospiraceae (Rosa *et al.*, 2005). *Borrelia* are classified as part of the Spirochaetaceae family along with *Treponema pallidum*, the causative agent of syphilis and yaws (Fraser *et al.*, 1998, Paster and Dewhirst, 2000) Other disease causing bacteria are present in the other two families; Leptospiraceae contains *Leptospira*, which is responsible for leptospirosis, while Brachyspiraceae encompasses *B. aalborgi* and *B. pilosicoli* which causes intestinal spirochaetosis. (Peruzzi *et al.*, 2007)



**Figure 1.3. 16s rRNA Phylogenetic tree of Spirochaetes.** Phylogenetic tree adapted from (Paster and Dewhirst, 2000). *B. burgdorferi* s.s. are closely related to *Cristispira*, a Spirochaete found in the digestive tract of oysters (Paster and Dewhirst, 2000, Madigan and Martinko, 2006) and *Treponema* and *Spirochaeta*. *Brachyspira* and *Leptospira* are further removed from *Borrelia*.

*Borrelia* are microaerophilic obligate parasites, requiring an environment with lower than atmospheric oxygen levels and are therefore unable to survive outside a host (Fraser *et al.*, 1997, Paster and Dewhirst, 2000, Neelakanta *et al.*, 2007). In addition, *Borrelia* have a minimal genome (Moran, 2002) and lack a respiratory electron transport chain, thus relying on substrate level phosphorylation mechanisms. *Borrelia* are therefore dependent on either the mammalian or arthropod host for their survival, nutritional requirements and transmission (Fraser *et al.*, 1997, Steere, 2001). The Spirochaetes have evolved and adapted complex mechanisms for long term survival in each host environment (Purser *et al.*, 2003). Natural transmission of *Borrelia* usually involves warm blooded mammals, birds and rodents as

hosts, enabling *Borrelia* to spread to and infect the non-infected tick, furthering the bacterial enzootic cycle (Alitalo, 2004).

The three species that will be focused on throughout this research are the major species responsible for causing Lyme Disease in humans: *B. afzelii*, *B. burgdorferi* s.s. and *B. garinii* (Kurtenbach *et al.*, 2006). *B. afzelii* and *B. garinii* are most common in Europe and Asia, although *B. burgdorferi* s.s. is present (Bunikis *et al.*, 1995, Wang *et al.*, 1999, Rosa *et al.*, 2005, Casjens *et al.*, 2011) (Figure 1.4). Whilst all three species cause Lyme Disease, the latter stage symptoms that occur are dependent on the species of the infecting Spirochaete (section 1.1.1) (Kalish *et al.*, 1993, Pal *et al.*, 2000). The difference in symptoms is potentially due to the different species of *Borrelia* evolving in different climates and environments, using different mammalian species with varying immune systems as host reservoirs.



**Figure 1.4. Geographical distribution of various *Borrelia* species.** The above figure shows the distribution of the major disease causing species of *Borrelia* in North America and Europe/Russia. Solid red circles show areas where the presence of *B. burgdorferi* s.s. has been confirmed, the green dotted line encircled area shows the region where *B. afzelii* has been found and the blue dash encircled area highlights confirmed *B. garinii* presence. Data adapted from (Kurtenbach *et al.*, 2006), map licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.

### 1.2.1 The genetics of *Borrelia*

Since the first sequencing of a whole *B. burgdorferi* s.s. genome (strain B31) (Fraser *et al.*, 1997), the whole-genomes of fourteen strains of *B. burgdorferi* s.s. (Casjens *et al.*, 2000, Schutzer *et al.*, 2011), two strains of *B. afzelii* (PKo and ACA-1) and two *B. garinii* strains (PBr and Far04) have been sequenced (Casjens *et al.*, 2011). *B. burgdorferi* s.s. (B31) was the third ever bacterial genome to be sequenced following *H. influenza* and *M. genitalium* (Fleischmann *et al.*, 1995, Fraser *et al.*, 1995).

*Borrelial* DNA is composed of a linear chromosome, and both closed hairpin and circular plasmids, a phenomenon which is unique among bacteria (Barbour, 1993, Fraser *et al.*, 1997,

Krupka *et al.*, 2007, Lescot *et al.*, 2008). The genome of *Borrelia* is unusually A-T-rich, approximately 71 % of the entire genome is composed of either adenine or thymine (Hyde and Johnson, 1984, Fraser *et al.*, 1997, Howe *et al.*, 2003). As *Borrelia* are dependent on the host for survival and nutrients, it has been proposed that an A-T rich genome is a result of a trade-off to incorporate ATP and TTP nucleotides rather than the more energetically expensive and less readily available GTP and CTP counterparts, due to the lower energy cost of producing ATP (Rocha and Danchin, 2002). It is also speculated that G-C content is related to phylogeny; 5s rRNA analysis suggested A-T/G-C pressure acts on the whole genome during evolution (Muto and Osawa, 1987), potentially as a result of the easier availability of ATP rather than GTP, which results in modification of components of DNA synthesis biased mutation rate towards A-T or methylation and deamination of DNA bases results in accumulation of A-T pairs (Muto and Osawa, 1987).

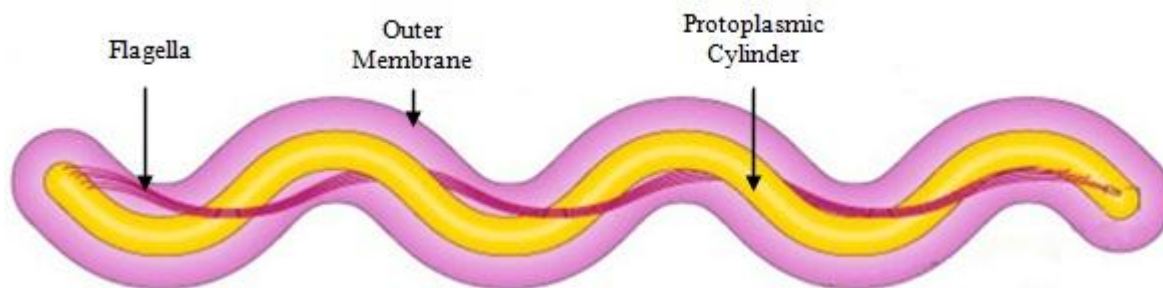
The A-T content of genes is up to 5 % less than that of the overall genomic content, but is still unusually high (Howe *et al.*, 2003). The genome of *B. burgdorferi* s.s. is 1.44 Mbp in length (Madigan and Martinko, 2006), Approximately 93 % of the *Borrelial* chromosome codes for essential proteins such as DNA replication, transcription, translation, solute transport and energy metabolism, but very few homologues to other eubacterial regulatory proteins have been identified in the *Borrelial* genome (Fraser *et al.*, 1997).

*Borrelia* have the largest number of known extrachromosomal DNA (ecDNA) fragments of any bacteria. The strain *B. burgdorferi* B31 has a total of 21 extrachromosomal elements ranging in size from 9 to 56 Kbp comprised of - 9 circular and 12 linear plasmids with covalently closed hairpin telomeres (Casjens *et al.*, 2000, Purser *et al.*, 2003, Rosa *et al.*, 2005). Only 71 % of the DNA encodes for putative genes, due to the presence of intergenic regions and a significant number of pseudo-genes in a state of decay with frameshifts and stop-codons (Fraser *et al.*, 1997). The *B. afzelii* and *B. garinii* strains sequenced in 2011 (Casjens *et al.*, 2011) established that the number of plasmids present in those strains varied from 7 to 17. Only approximately 8 % of the putative genes encoded in ecDNA were shown to be similar to genes in other bacteria, however none of these similar genes were virulence factors. This suggests that either the Spirochaete has evolved with different virulence and pathogenic proteins, has remote homologues, or both (Casjens *et al.*, 2000). Plasmids appear to be consistent in a variety of *Borrelial* species and strains, which implies that they are important for survival, infectivity and propagation (Purser *et al.*, 2003). Sixty three putative

membrane lipoproteins have been identified as being plasmid-encoded, providing evidence for the differential expression of proteins in various hosts (Fraser *et al.*, 1997). A study of *B. burgdorferi* s.s. indicates that when cultured in-vitro the bacteria are prone to losing plasmids. This has been shown to have minimal effect on propagation, but does result in a loss of infectivity in mice (Ristow *et al.*, 2012) as a significant number of outer membrane proteins (Steere, 2001), several virulence and infective proteins (PncA, OspA-D, DbpA and DbpB, VlsE) are plasmid-encoded (Comstock *et al.*, 1993, Fraser *et al.*, 1997, Roberts *et al.*, 1998, Kumaran *et al.*, 2001, Hübner *et al.*, 2001, Purser *et al.*, 2003, Stewart *et al.*, 2005). For example, the removal of plasmid lp25 encoding a cytoplasmic nicotinamidase (PncA) was sufficient to knock out *Borrelial* infectivity in rats, replacement of the gene via transformation was sufficient to restore infectivity (Purser *et al.*, 2003).

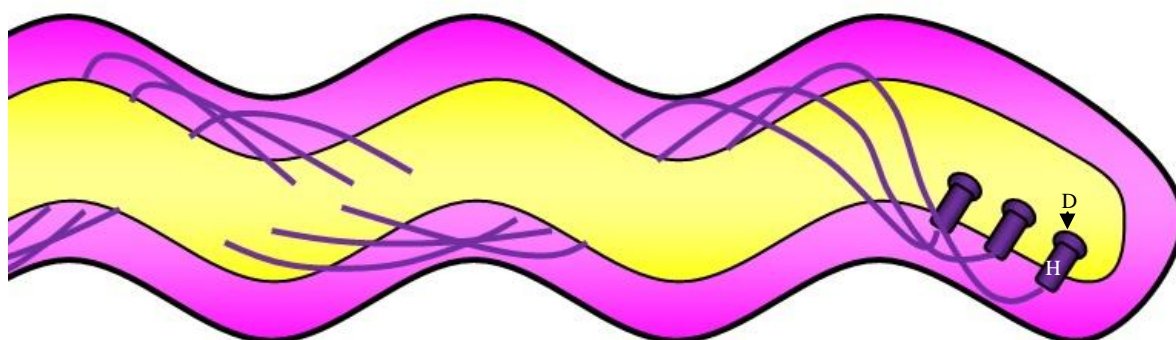
### **1.2.2 The ultrastructure of *Borrelia***

*Borrelia* are tightly coiled and highly motile Spirochaete bacteria (Figure 1.5) (Madigan and Martinko, 2006). *Borrelia* were initially assumed to be Gram-negative due to the presence of both an outer and inner (cytoplasmic) membrane. However as techniques advanced and more detailed analyses of the genetics, molecular architecture and biochemical workings were undertaken on *Borrelia* and other Spirochaetes (including *Spirochaeta*, *Leptospira* and *Treponema*) a clear difference was seen and a new eubacterial phylum was created (Paster *et al.*, 1991). Unlike Gram-negative bacteria, *Borrelia* and *Treponema* lack lipopolysaccharide (LPS) on the outer surface of the membrane (Takayama *et al.*, 1987), although *Leptospira* and *Brachyspira* do synthesise LPS (Saier, 2000). From the phylogenetic tree (Figure 1.3) it is apparent that *Borrelia* and *Treponema* are more closely related to each other than *Leptospira* and *Brachyspira*. Another major difference is the OM proteome, Gram-negative bacteria contain a high number of transmembrane proteins, with few exposed lipoproteins, in *Borrelia* the opposite is true (Cox *et al.*, 1996). *Borrelia* vary in length from 10 – 30 µm in length and approximately 0.2 to 0.3 µm in diameter (Krupka *et al.*, 2007).

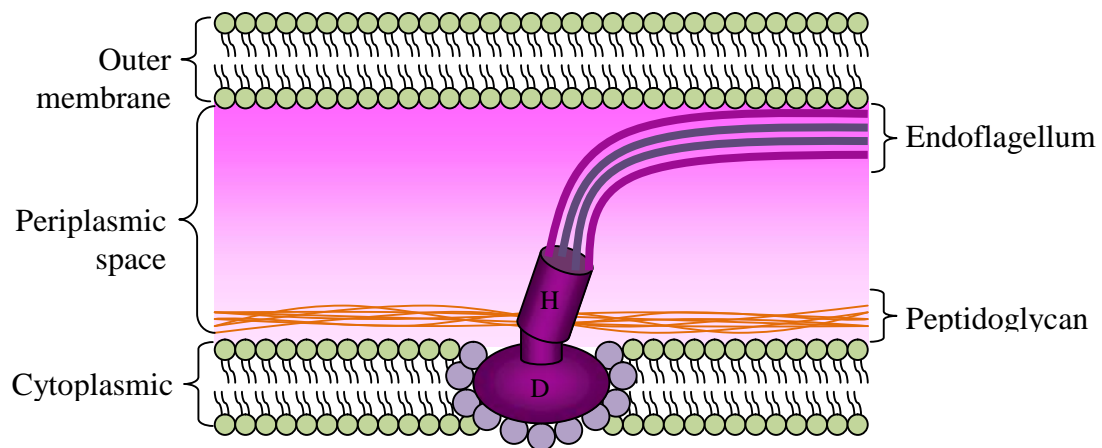


**Figure 1.5. Basic diagram of the ultrastructure of a *Borrelia* Spirochaete** Diagram adapted from (Rosa *et al.*, 2005). The bacterium has both an outer and a cytoplasmic membrane, with multiple endoflagellum present within the periplasmic space between. All the other cellular components such as DNA are contained in the protoplasmic cylinder.

The motility of Spirochaetes differs from that of bacteria with extracellular flagella; one end of each endoflagella is attached to the cytoplasmic membrane of the protoplasmic cylinder towards the terminus of the Spirochaete, extending along the Spirochaete and overlapping with endoflagella attached at the other terminus (Madigan and Martinko, 2006, Charon *et al.*, 2009). The endoflagella are anchored at one terminus into the cytoplasmic membrane as shown below (Figure 1.6 & Figure 1.7). *Borrelia* contain between 7 and 11 endoflagella per origin (Hovind-Hougen, 1984, Motaleb *et al.*, 2000). The flagella also play a vital role in the skeletal structure of *Borrelia* as flagella deficient *Borrelia* are rod-shaped, this is unusual as bacterial shape is normally dictated by the layer of peptidoglycan (Motaleb *et al.*, 2000, Charon *et al.*, 2009).

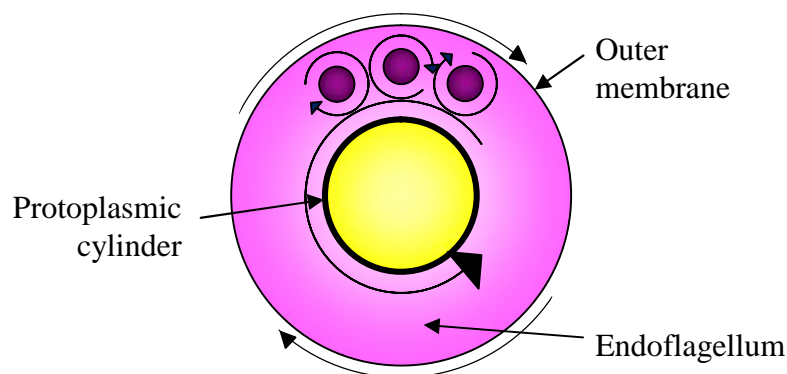


**Figure 1.6. Basic diagram showing the anchoring points of endoflagella.** Highlighted are the discs and hooks (D and H respectively) present at each terminus of the flagella. Only one end of each endoflagella is anchored in the protoplasmic cylinder and endoflagella from each terminus of the bacteria overlap in the centre. Diagram based on an SEM image from (Hovind-Hougen, 1984).



**Figure 1.7. Anchoring of endoflagella in the cytoplasmic membrane of *Borrelia*** The disc and hook that anchor the endoflagella into the membrane are labelled D and H respectively. Diagram based on (Rosa *et al.*, 2005). Rotation of the flagella is a result of proton transport-coupled rotation, which causes an element of the disc to spin around a central core (Berg *et al.*, 2002). The rotation is capable of changing direction rapidly, which is useful for chemotaxis (Berg *et al.*, 2002).

Co-ordinated rotation of the endoflagella in the same direction causes the protoplasmic cylinder to rotate in the opposite direction while the outer membrane rotates in the same direction as the endoflagella (Figure 1.8). This places torsion on the cell, conferring motility through flexing motions; this is useful in highly viscous environments such as cell cytoplasm and helps with transition between tissues from initial deposition in the skin. The Spirochaete can reach speeds of  $4 \mu\text{m}/\text{sec}^{-1}$ , fast enough to evade phagocytes (Madigan and Martinko, 2006, Radolf *et al.*, 2012). Some components of bacterial flagella (principally flagellin) are known to stimulate a pro-inflammatory immune response in hosts (Zgair, 2012). The location of *Borrelial* flagella within the protoplasmic cylinder means that the antigenic components of flagella are not normally presented to the immune system.

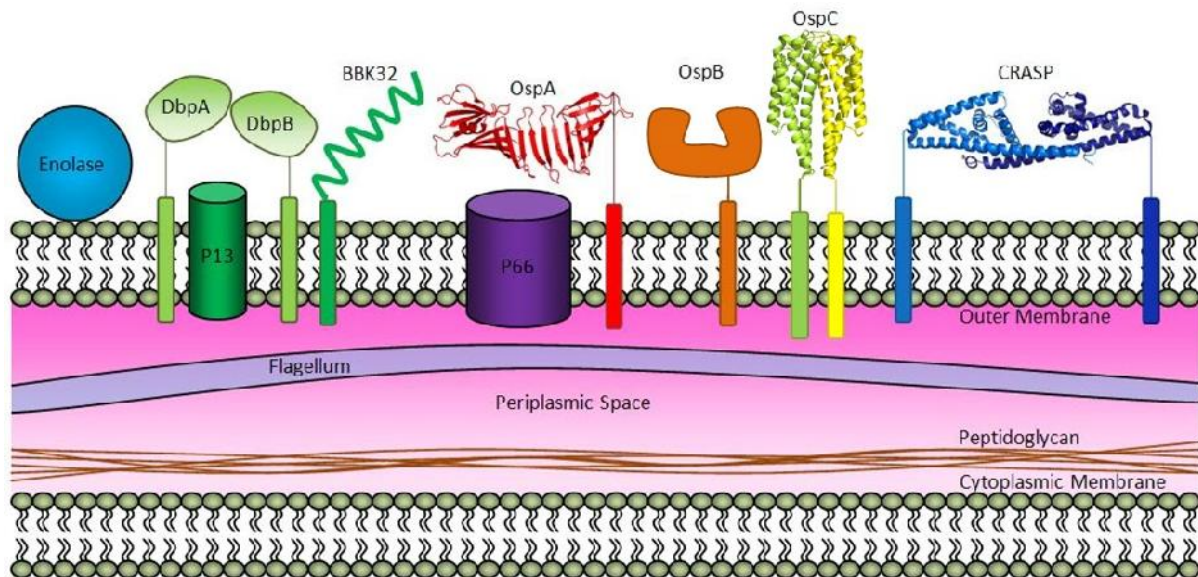


**Figure 1.8. Cross section of a *Borrelia* bacterium.** Based on (Madigan and Martinko, 2006), as the endoflagella rotate in one direction, the protoplasmic cylinder rotates in the opposite direction. As in a system of cogs, the outer membrane rotates in the same direction as the endoflagella. The endoflagella are arranged side-by-side as in a ribbon, rather than bundled together (Charon *et al.*, 2009).



### 1.2.3 The outer membrane of *Borrelia*

Expression of proteins in the outer membrane of *Borrelia* has been shown to vary according to many different factors, including: the host species the Spirochaete is occupying, temperature, the location within the host and the stage of infection (Fraser *et al.*, 1997, Pal *et al.*, 2000, Babb *et al.*, 2001, Pal *et al.*, 2004b). This is vital to the Spirochaete evading the immune system responses of the hosts and sustaining an infection. Several types of protein of known function have been shown to be present on the outer membrane of *Borrelia* at one time or another (Figure 1.9); the most prominent of which are a variety of lipoproteins including decorin binding proteins, outer surface proteins, CRASPs and complement binding proteins; an enolase and porins (described in section 1.2.4.5) (Skare *et al.*, 1997, Cassatt *et al.*, 1998, Schwan and Piesman, 2000, Babb *et al.*, 2001, Fikrig *et al.*, 2004, Pal *et al.*, 2004b, Brooks *et al.*, 2006, Bárcena-Uribarri *et al.*, 2010).



**Figure 1.9. Schematic representation of known lipoproteins and outer membrane proteins in *Borrelial* membranes.** Proteins of known structure are shown in the cartoon form, whilst proteins of unknown structure are shown as solid shapes. The solid rectangles are representative of lipid anchors. Not all proteins in the diagram are upregulated at the same time, for example OspA and OspC (Schwan and Piesman, 2000). OspB is represented as a protein with a cleft in the middle as only the structure of the C-terminus has been solved (Becker *et al.*, 2005). There are three more known Osp (D-F) but these are not as thoroughly characterised as Osp A-C and so are not included.

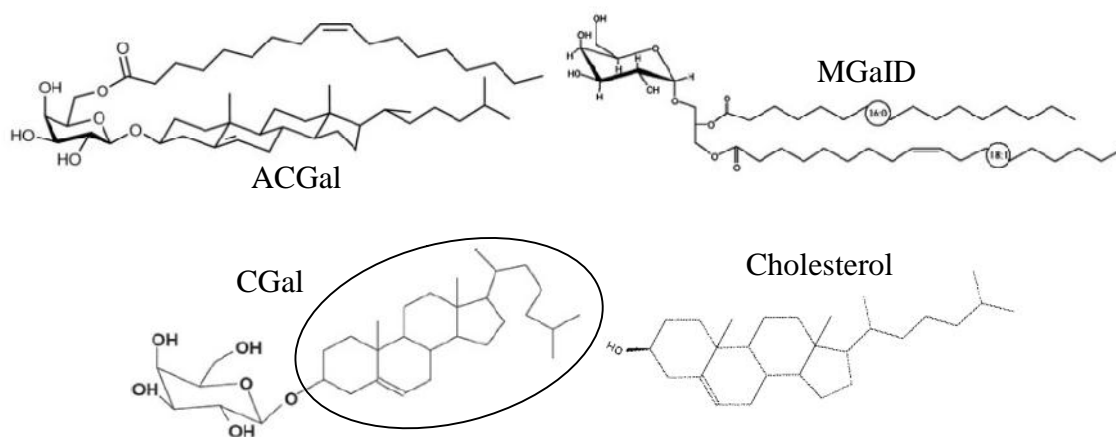
As mentioned previously (section 1.2.2), unlike Gram-negative bacteria, lipopolysaccharide (LPS) is not expressed by *Borrelia*; it has been speculated that this is part of the Spirochaetes immune system evasion. LPS is readily detectable and highly antigenic (Takayama *et al.*, 1987, Seemanapalli *et al.*, 2010). *Borrelia* express the largest number of lipoproteins of any bacteria, with over 175 known (approximately 10 % of the genome codes for lipoproteins)



(Fraser *et al.*, 1997, Eicken *et al.*, 2001, Templeton, 2004, Yang *et al.*, 2005) Spirochaetal lipoproteins are anchored to the outer membrane through lipidation (Seemanapalli *et al.*, 2010) and are activators of the innate immune response in mammals. These lipoproteins are differentially expressed depending on the environmental conditions in which the Spirochaete currently resides (Bunikis *et al.*, 1995).

### 1.2.3.1 Lipid rafts in the OM of *Borrelia*

Recently the outer membrane of *Borrelia* have been shown to contain free cholesterol, cholesterol esters and three types of glycolipid (Figure 1.10); 6-O-acylated -D-galactopyranoside (ACGal), cholesteryl- -D-galacto-pyranoside (CGal) and mono- -D-galactosyldiacylglycerol (MGaID) (Ben-Menachem *et al.*, 2003, Schröder *et al.*, 2003, Stübs *et al.*, 2009, LaRocca *et al.*, 2010).

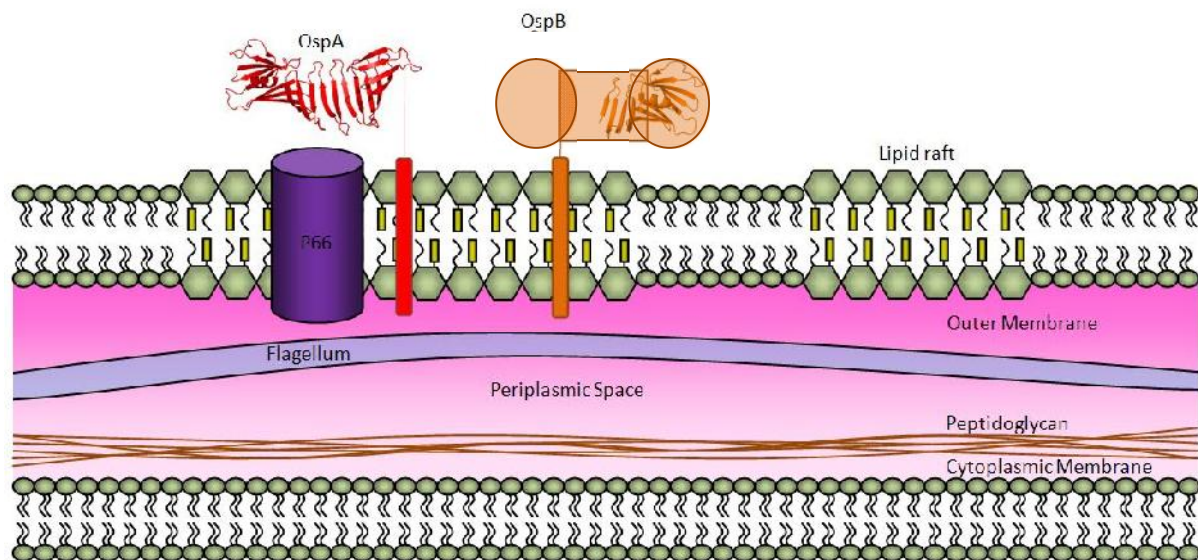


**Figure 1.10. Glycolipids present in the OM of *Borrelia*.** ACGal (Stübs *et al.*, 2009); MGaID (Ben-Menachem *et al.*, 2003); CGal (Dorasamy *et al.*, 2012). Both ACGal and MGaID have a similar length alkyl chain with a double bond present and have antigenic properties (Crowley *et al.*, 2012). CGal contains a cholesterol group (highlighted by the black circle).

This is unusual in that aside from a small number of prokaryotes; *Mycoplasma*, *Helicobacter*, *Ehrlichia*, *Anaplasma* and *Brachyspira*, only eukaryotic cells have previously been shown to contain free cholesterol (LaRocca *et al.*, 2010). Glycolipids and cholesterol make up a significant volume of the outer membrane of the *Borrelia* (LaRocca *et al.*, 2010); this varies between 56.6 and 51.8 % of the total lipids depending upon the species (Stübs *et al.*, 2009). Despite containing cholesterol and cholesterol glycolipids, *Borrelia* lack the ability to synthesise cholesterol or indeed any long chain fatty acids, both saturated and unsaturated (Crowley *et al.*, 2012). It has been shown that two-way lipid exchange between the Spirochaete and host occurs, resulting in acquisition of cholesterol by *Borrelia* and transfer of

ACGal and CGal to the outer membrane of mammalian host cells (Crowley *et al.*, 2012). Transfer potentially occurs either through direct *Borrelial* contact with host cells or through Spirochaetal outer membrane vesicle contact; the presence of *Borrelial* antigens on host cells is speculated to contribute to inflammation, arthritis and targeting of host cells by the immune system (Crowley *et al.*, 2012).

It has been shown that cholesterol-rich lipid rafts (**Error! Reference source not found.**) similar to those found in eukaryotic bacteria are formed in the OM of *Borrelia* (not previously seen before in prokaryotes) and that OspA, OspB and P66 associate with these rafts (Figure 1.11). The sizes of rafts vary with temperature changes, potentially aiding in dissemination (LaRocca *et al.*, 2010). Other tick-borne pathogens *E. chaffeensis* and *A. phagocytophilum* contain free cholesterol and lipid rafts.



**Figure 1.11. Cholesterol rich lipid rafts in the outer membrane of *Borrelia*.** Lipid rafts are ordered and rigid regions of the OM rich in cholesterol (LaRocca *et al.*, 2010). The green hexagons in the OM represent glycolipids, yellow rectangles represent cholesterol and the curved lines attached to green hexagons represent palmitoyl. As only half of the structure of OspB is known the overall structure is represented with basic shapes and the solved structure overlain. The cleft in the structure of OspB has been shown to weakly bind cholesterol glycolipids, potentially aiding the formation of lipid rafts in *Borrelia* (LaRocca *et al.*, 2010).

## 1.2.4 Immune evasion and infectivity of *Borrelia*

Upon initial transmission to a mammalian host, *Borrelia* remain localised around the site of the tick bite and the infection can be cleared by a course of antibiotics. If left untreated however, the Spirochaete can disseminate to tissues, organs and joints, leading to Lyme disease, neuroborreliosis and arthritis (Burkot *et al.*, 1994). *Borrelia* numbers in the tick

midgut have been shown to increase markedly after the completion of a bloodmeal, and decrease before the arthropod moults to the next stage in its life cycle (Burkot *et al.*, 1994). The survival of *Borrelia* is challenging as the bacteria has to be able to exist in two distinct host environments, in addition to facing changes within each environment; including the influx of blood into the midgut and physiological changes caused by tick moulting as well as the immune responses of both hosts (Schwan and Piesman, 2000).

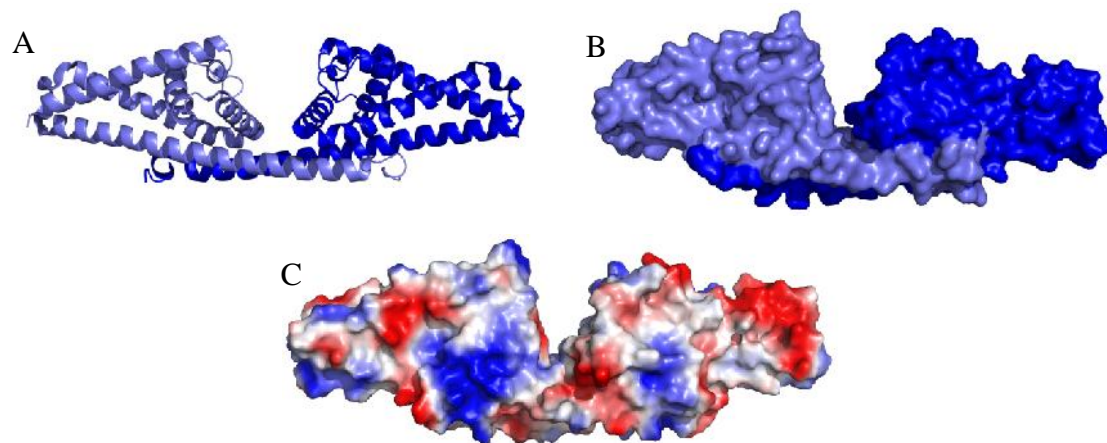
*Borrelial* outer membrane proteins are of vital importance to both the bacteria and the infected host; they provide both the bacteria with means to evade the immune system and survive in the host, while at the same time providing the host with targets for bacterial removal. The two most studied groups of surface exposed *Borrelial* lipoproteins are *B. burgdorferi* complement regulator acquiring surface proteins (BbCRASPs) and lipid-anchored outer surface proteins (Osps). CRASPs have been shown to bind Factor H and related proteins, negating activation of the alternative pathway of the complement immune system (Cordes *et al.*, 2005, Cordes *et al.*, 2006, Kraiczy *et al.*, 2006, Rossman *et al.*, 2006, Kraiczy *et al.*, 2009). Osps found in *Borrelia* are genetically and antigenically polymorphic and have been shown to trigger immune responses in mammals (Bunikis *et al.*, 1995). Six Osps named A through F are known (Lam *et al.*, 1994, Bu *et al.*, 1998, Babb *et al.*, 2001, Becker *et al.*, 2005, Hovis *et al.*, 2006, Li *et al.*, 2007). Osps aid the Spirochaete in transmission to, and establishment of the initial infection in mammals and persistence in arthropod vectors (Cassatt *et al.*, 1998, Burkot *et al.*, 1994, Pal *et al.*, 2000, Pal *et al.*, 2004a, Pal *et al.*, 2004b). It has also been shown that Osps can be localised on the cytoplasmic membrane (CM), protruding into periplasmic space and it is speculated that the CM can be used as a 'holding area' (Cox *et al.*, 1996) meaning that the proteins can rapidly be transported to the outer membrane in response to specific environmental triggers.

#### **1.2.4.1 BbCRASPs**

*B. burgdorferi* s.s. Complement Regulator Acquiring Surface Proteins (BbCRASPs) are plasmid encoded lipid-anchored proteins expressed on the surface of the outer membrane that play a role in *Borrelial* complement evasion (Coleman *et al.*, 2008). Five have been identified in both *B. burgdorferi* s.s. and *B. afzelii* (Hovis *et al.*, 2006, Rossman *et al.*, 2006, Kraiczy *et al.*, 2001a, Brooks *et al.*, 2005). CRASPs bind mammalian complement regulating proteins Factor H, Factor H-like protein (FHL-1) and Factor H-related protein (FHR-1), which are

responsible for the recognition of host cells and mediation of the alternative pathway of the complement immune system (Pangburn, 2000). The binding of these proteins prevents activation of the alternative pathway of the host's immune response and aids *Borrelia* survival and dissemination (Kraiczy *et al.*, 2001b).

The structure of BbCRASP-1 has been solved by X-Ray crystallography (Figure 1.12) (Cordes *et al.*, 2005); each BbCRASP-1 contains five  $\alpha$ -helices and it was found that BbCRASP-1 forms a biologically relevant homodimer on the Spirochaete surface (Cordes *et al.*, 2006). It is known to interact with short consensus repeat (SCR) domains of Factor H and FHL-1 (Cordes *et al.*, 2005, Hovis *et al.*, 2006, Cordes *et al.*, 2006, Kraiczy *et al.*, 2009). The C-terminus of each protein in the dimer are buried at the centre of the overall structure and are involved in fH and fHL-1 binding (Kraiczy *et al.*, 2006). BbCRASP-1 knockout *Borrelia* are vulnerable to killing by complement, suggesting that BbCRASP-1 plays a role in evasion and infection of mammals. There is significant evidence showing the expression of BbCRASP-1 on the surface of *Borrelia* when the Spirochaete is present in mammals, and in the transition between mammals and arthropods (Rossman *et al.*, 2006, Hovis *et al.*, 2006, Kraiczy *et al.*, 2009). The loss of BbCRASP-1 has been shown to make the Spirochaete susceptible to complement mediated killing by mammalian sera (Coleman *et al.*, 2008).



**Figure 1.12. Crystal structure of BbCRASP-1.** A 27.5 kDa protein expressed on the surface of *Borrelia* (Kraiczy *et al.*, 2001a). **A)**, **B)** and **C)** are the cartoon, surface and surface electrostatic potential representations respectively. The structure was solved at a resolution of 2.7 Å (RCSB PDB accession 1W33) (Cordes *et al.*, 2005). Clearly visible in the centre of the dimer is a cleft, containing the binding sites predicted to bind SCR domains of human factor H (Kraiczy *et al.*, 2009). The cleft is approximately the correct size to accommodate SCR domains lying along their axes (Cordes *et al.*, 2006).

BbCRASP-2 is not expressed in by the Spirochaete when in arthropod hosts but is expressed consistently in infection of mammals (Coleman *et al.*, 2008, Kraiczy *et al.*, 2009).

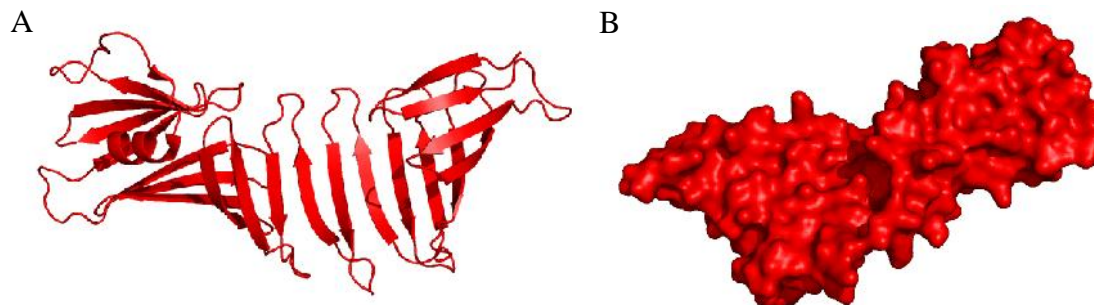
BbCRASP-2, like BbCRASP-1 is known to bind Factor H and FHL-1 (Kraiczy *et al.*, 2006, Coleman *et al.*, 2008, Kraiczy *et al.*, 2008), though studies have shown that BbCRASP-2 deficient *B. burgdorferi* s.s. have no loss of infectivity or diminished survival rates. It is therefore likely that BbCRASP-2 is not essential for effective immune evasion and colonisation of mammalian hosts, but does, however, play a role (Coleman *et al.*, 2008). This is supported by evidence that BbCRASPs 1 and 2 bind factor H and factor H-like protein through SCR-7 (Reuter *et al.*, 2010).

BbCRASPs 3, 4 and 5 all have very similar amino-acid sequences and are distinctly different from BbCRASPs 1 and 2, hence being known collectively as OspE (Coleman *et al.*, 2008) and OspE related proteins (Erps). (McDowell *et al.*, 2003, Kraiczy *et al.*, 2006). OspE are lipoproteins (Lam *et al.*, 1994) are expressed during the initial stages of mammalian infection (Babb *et al.*, 2001) and have been shown to bind complement regulating Factor H and Factor H-related protein (FHR-1). (Hovis *et al.*, 2006, Coleman *et al.*, 2008, Kraiczy *et al.*, 2009). The structures have not yet been solved.

#### **1.2.4.2 Outer Surface Protein A**

OspA is a dumbbell shaped immunogenic lipoprotein anchored in the outer membrane of *Borrelia* by a lipidated N-terminal cysteine (Bu *et al.*, 1998, Huang *et al.*, 1998). It is composed of 273 amino-acids (including signal sequence) and has a molecular weight of 28 kDa (Huang *et al.*, 1998). The protein forms a twenty one stranded anti-parallel  $\beta$ -sheet with a single  $\alpha$ -helix at the C-terminus (Figure 1.13) (Huang *et al.*, 2001). Each end of the  $\beta$ -sheet forms a globular domain, linked by a non-globular region in the middle (Bu *et al.*, 1998). The central  $\beta$ -sheet is exposed on both sides to any external influences on the protein, including any from a solvent during crystallisation. This tertiary structure of OspA is unusual in regards to the exposure of polar residues in the central sheet to solvent, in general proteins containing anti-parallel  $\beta$ -sheets are amphipathic, with the hydrophobic amino-acids sheltered in a hydrophobic core (Li *et al.*, 1997, Bu *et al.*, 1998), however it appears that the hydrophobic residues in the central  $\beta$ -sheet are protected by an abundance of amino-acids with long side chains, increasing the thickness of the  $\beta$ -sheet (Huang *et al.*, 2001). SAXS (Bu *et al.*, 1998) solution NMR studies (Huang *et al.*, 1998, Huang *et al.*, 2001) have shown that OspA exists in solution as a monomer, and is in agreement with the solved structure that the protein consists of two globular domains linked by a stable central  $\beta$ -sheet, hence does not have a

hydrophobic core. This is highly unusual for proteins, another protein known to lack a hydrophobic core is SasG, a protein from *S. aureus* known to be involved in the formation of biofilms (Geoghegan *et al.*, 2010)



**Figure 1.13. Crystal structures of OspA from *B. burgdorferi* s.s.** Data retrieved from the RCSB Protein Data Bank (Accession 2G8C). **A)** Cartoon structure showing  $\beta$ -sheets and the lone  $\alpha$ -helix. **B)** Surface representation of OspA. X-Ray diffraction data was collected with a resolution of 1.15 Å. The protein was crystallised in complex with mouse monoclonal Fab fragment (not shown) (Makabe *et al.*, 2006). Images of the data created using PyMol 1.3, (Schrödinger, 2010).

OspA is abundantly expressed on the surface of *Borrelia* while the Spirochaete is present in the midgut of the tick (Li *et al.*, 1997, Pal *et al.*, 2000, Schwan and Piesman, 2000, Purser *et al.*, 2003, Becker *et al.*, 2005, Brooks *et al.*, 2006). It has been shown that OspA is upregulated when entering the midgut of an uninfected tick (Pal *et al.*, 2000, Fikrig *et al.*, 2004) and downregulated upon the tick partaking in a blood-meal, prior to *Borrelial* transmission to a mammalian host (Burkot *et al.*, 1994, Schwan *et al.*, 1995, Cassatt *et al.*, 1998, Pal *et al.*, 2000, Schwan and Piesman, 2000, Liang *et al.*, 2004). The peptide sequence and antigenic reactions of OspA are consistent between *Borrelial* species (Becker *et al.*, 2005).

OspA facilitates *Borrelia* adherence to, and colonisation of the midgut of ticks prior to feeding on a mammalian host (Pal *et al.*, 2000). It was shown that recombinant OspA bound specifically to tick gut extract and other OspA molecules, whereas OspC and BSA used as control proteins did not bind either OspA or the gut extract (Pal *et al.*, 2000). It was found that a 55 kDa tick receptor protein for OspA (TROSPA) was responsible for binding OspA and was localised in the gut, and that the presence of *Borrelia* significantly increased the expression of TROSPA by the tick (Pal *et al.*, 2004a, Neelakanta *et al.*, 2007). OspA-OspA and binding appears to facilitate binding of *Borrelia* to the midgut of ticks; addition of excess OspA to a mixture of *Borrelia* and tick gut extract was shown to enhance bacterial adhesion.

It is also feasible that OspA-OspA binding could also facilitate the binding of *Borrelia* to each other, aiding the establishment of *B. burgdorferi* s.s. in the tick midgut (Pal *et al.*, 2000). Both deficiencies in TROSPA in the tick, and OspA in the Spirochaete resulted in reduced colonisation (Pal *et al.*, 2000, Pal *et al.*, 2004a).

Downregulation of OspA by the Spirochaete, and TROSPA by the tick during a bloodmeal potentially aids the Spirochaete in detaching from the tick midgut for transmission to the mammalian host (Pal *et al.*, 2004a). This would prevent the bacteria from binding to each other and facilitate easier translocation to the salivary glands and therefore transmission to the mammalian host (Burkot *et al.*, 1994, Cassatt *et al.*, 1998, Pal *et al.*, 2000). OspA is not expressed on the surface of *Borrelia* when the Spirochaete is present in mice; this is likely due to the protein having an function in ticks but not mammals, hence downregulation to prevent immune response targeting (Liang *et al.*, 2004).

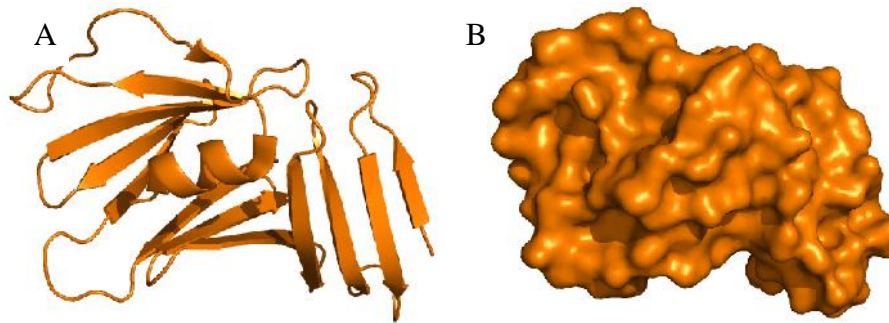
It has recently been shown that *B. burgdorferi* s.s. form biofilms (Sapi *et al.*, 2012), which may be a factor in the development of Lyme arthritis. It is possible that OspA-OspA binding may play an important role in facilitating *Borrelia-Borrelia* binding and formation of biofilms. OspA has been shown to be continually repressed in mammalian host to prevent host antibodies binding and eliciting an immune response (Schwan *et al.*, 1995, Pal *et al.*, 2000, Schwan and Piesman, 2000), however evidence is present that OspA can be expressed in the mammalian host; antibodies for OspA have been identified in the cerebrospinal fluid of patients with neuroborreliosis (Liang *et al.*, 2004) and joints of patients with Lyme arthritis (Kalish *et al.*, 1993). The latter study monitored immune responses to fifteen patients infected with *Borrelia* over a period of months and years. Antibodies for both OspA and OspB were found in eleven patients just prior to the onset of arthritis, having not been present in the months and weeks before (Kalish *et al.*, 1993, Pal *et al.*, 2000), antibodies for OspA and OspB were found in the remaining four patients as the symptoms progressed. This may be another indicator of OspA importance in biofilm formation, and of biofilm importance in arthritis.

#### **1.2.4.3 Outer Surface Protein B**

OspB is a 296 residue, 31 kDa protein lipid-anchored protein present on the outer membrane of *Borrelia* when in the arthropod midgut (Becker *et al.*, 2005) that is closely related in both sequence and structure to OspA (Neelakanta *et al.*, 2007). The overall shape of OspB is

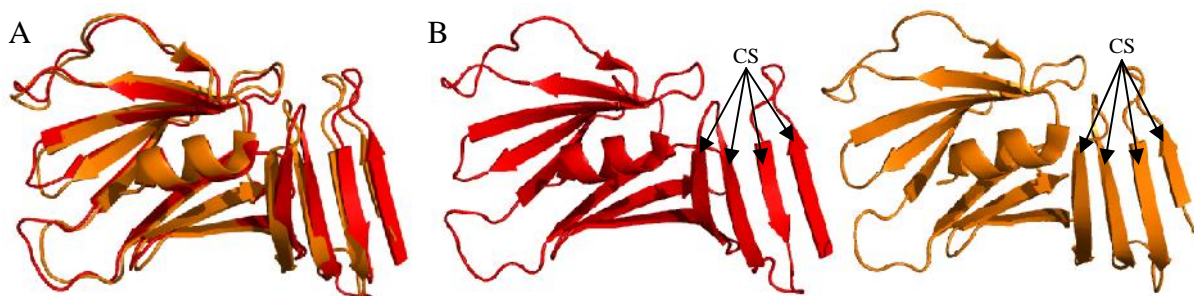


unknown, however the structure of the 140 residue C-terminal domain has been solved (Figure 1.14) (Becker *et al.*, 2005). The C-terminus of OspB is composed of twelve anti-parallel  $\beta$ -strands and one  $\alpha$ -helix and has been shown by X-ray crystallography to adopt the same configuration as that of OspA; overlaying very well (Figure 1.15 A) (Becker *et al.*, 2005).



**Figure 1.14. Crystal structure of the 16.5 kDa C terminus of OspB from *B. burgdorferi* s.s.** (Becker *et al.*, 2005) **A)** Cartoon structure showing the  $\beta$ -sheets and  $\alpha$ -helix present, **B)** Representation of the surface structure. The peptide chain extends from residue 157 to 296 and the structure solved at a resolution of 2.0 Å. Data retrieved from the RSCB Protein Data Bank (accession 1P4P). Images produced using PyMol 1.3 (Schrödinger, 2010).

Due to the similarities between the C-terminal domains of OspA and OspB (Figure 1.15 A) it is reasonable to assume that OspB also forms a globular domain at the C-terminal domain. In addition, the four  $\beta$ -sheets highlighted in OspB (Figure 1.15 B) are positioned in an identical manner to those in OspA, and almost certainly contributes to a central stable  $\beta$ -sheet as in OspA, meaning that the protein also lacks a hydrophobic core. OspB is a longer protein than OspA, having a 23 amino-acid insertion towards the N-terminus located before where the first  $\beta$ -strand would be if OspB maintains the structural similarities to OspA (Becker *et al.*, 2005).



**Figure 1.15. Comparison of the C-termini of OspA and OspB.** **A)** Overlay of OspA C-terminus with OspB C-terminus. **B)** OspA and OspB C-termini side by side comparison. OspA (PDB accession: 2G8C) is red, OspB (PDB accession: 1RJL) is Orange (Becker *et al.*, 2005, Makabe *et al.*, 2006).  $\beta$ -strands denoted CS are known to form part of a central  $\beta$ -sheet linking two globular domains in OspA, and likely fulfil a similar function in OspB.

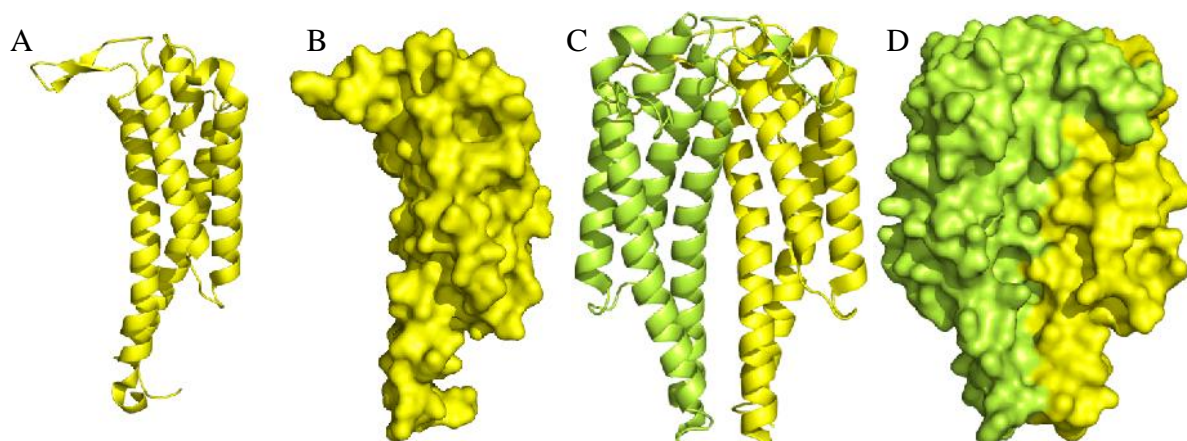


OspA and OspB genes are encoded by a single operon on an extrachromosomal plasmid (Fikrig *et al.*, 2004, Becker *et al.*, 2005, Neelakanta *et al.*, 2007) and it has been shown that OspA and OspB are transcribed together due to a shared promoter. OspB is primarily expressed while the Spirochaete is present in the midgut of unfed *Ixodid* ticks, but downregulated upon commencement of a bloodmeal (Becker *et al.*, 2005) and not expressed in mammals (Cassatt *et al.*, 1998). Like OspA, OspB has been shown to bind to the tick midgut (Fikrig *et al.*, 2004, Yang *et al.*, 2004, Neelakanta *et al.*, 2007). A study found that OspB specific antibodies inhibited the Spirochaete binding (Fikrig *et al.*, 2004), while another found that OspB deficient *B. burgdorferi* s.s. had an impaired ability to bind to the midgut of ticks upon commencement of the bloodmeal, but could still infect mice (Neelakanta *et al.*, 2007). These findings indicate that OspB is an important contributor to survival of *Borrelia* in ticks, but is downregulated in mammals to minimise potential antigenic surfaces presented to the mammalian host.

OspB, like OspA has also been implicated in Lyme arthritis (Kalish *et al.*, 1993). Prior to the onset of arthritis 73 % of patients developed a response to both OspA and OspB antibodies (Kalish *et al.*, 1993). In a different study, this provided further evidence for the presence of OspA and OspB prior to the onset of arthritis, however it also found that in mice OspB was expressed at considerable levels four months after developing arthritic symptoms (Liang *et al.*, 2004).

#### **1.2.4.4 Outer Surface Protein C**

OspC is a 23 kDa lipoprotein expressed on the surface of *Borrelia* (Babb *et al.*, 2001, Kumaran *et al.*, 2001, Pal *et al.*, 2004b) that contains four long anti parallel  $\alpha$ -helices and one short  $\alpha$ -helix (Figure 1.16 A & B), it is likely that the dimer shown in Figure 1.16 C and Figure 1.16 D is biologically relevant due to the formation of a four helix hydrophobic core, extending almost the whole length of the protein (Eicken *et al.*, 2001, Seemanapalli *et al.*, 2010). OspC has a strong negative electrostatic potential, potentially involved in binding charged host ligands – indicative of causing human disease; in addition, OspC is highly polymorphic (especially the surface exposed residues) making it difficult to design antibodies for (Wikel, 1996, Eicken *et al.*, 2001, Kumaran *et al.*, 2001).



**Figure 1.16. Crystal structure of OspC from *B. burgdorferi* s.s.** **A)** and **B)** - Cartoon and surface structures of monomeric OspC solved to a resolution of 2.5 Å (Eicken *et al.*, 2001). **C)** and **D)** cartoon and surface structures of dimeric OspC solved to a resolution of 2.5 Å (Kumaran *et al.*, 2001). X-Ray diffraction data was collected with a resolution of 2.5 Å. Images of monomeric OspC (A and B) and dimeric OspC (C and D) were produced using data obtained from the RSCB Protein Data Bank - accession numbers 1G5Z and 1F1M respectively (Eicken *et al.*, 2001, Kumaran *et al.*, 2001). Images produced using PyMol 1.3 (Schrödinger, 2010).

*Borrelia* OspC is vital for transmission of the Spirochaete from the midgut to the salivary glands of *Ixodes* ticks and for dissemination to mammalian hosts, playing an important role in establishment of infection and survival (Cassatt *et al.*, 1998, Eicken *et al.*, 2001, Hübner *et al.*, 2001, Pal *et al.*, 2004b, Ramamoorthi *et al.*, 2005, Yang *et al.*, 2005). OspC deficient *B. burgdorferi* s.s. have no infectivity (Seemanapalli *et al.*, 2010). OspC is upregulated in the salivary glands of arthropods (Schwan and Piesman, 2000, Pal *et al.*, 2004b), continually expressed in mammals throughout the early stages of infection (Babb *et al.*, 2001, Pal *et al.*, 2004b) and downregulated once the anti-OspC humoral response kicks in (Seemanapalli *et al.*, 2010). When *Borrelia* are present in the midgut of an unfed *Ixodes* tick, OspA and OspB are predominantly expressed (Liang *et al.*, 2004, Yang *et al.*, 2004), however upon the commencement of a bloodmeal and influx of mammalian blood the Spirochaete has been shown to rapidly upregulate OspC (Schwan and Piesman, 2000, Pal *et al.*, 2004b) and heavily downregulate OspA and OspB (Schwan *et al.*, 1988, Schwan *et al.*, 1995). This is triggered by the change in temperature and pH (Kumaran *et al.*, 2001). Upregulation of OspC assists the bacteria in translocation from the midgut of the tick to the salivary glands, and from there to the host (Eicken *et al.*, 2001). It was shown that OspC is capable of binding the salivary glands of ticks, and that OspC deficient bacteria were not capable of invading invade arthropod salivary glands; existence in tick midgut was unaffected (Pal *et al.*, 2004b). *Borrelia* in the initial stages of mammalian infection that are lacking OspC are swiftly cleared (Seemanapalli *et al.*, 2010).

OspC has been shown to bind tick salivary protein 15 (Salp15) both *in vivo* and *in vitro*. (Ramamoorthi *et al.*, 2005). The effect of Salp15 on *Borrelial* dissemination and infection of mice was studied and it was found that Salp15 binding significantly increased the survivability of *B. burgdorferi* s.s. in mice, along with protecting the Spirochaete from being by OspA antibodies (Ramamoorthi *et al.*, 2005). OspC antibodies have been detected in humans patients prior to onset of Lyme arthritis (Kalish *et al.*, 1993), this is not unexpected as OspC is an important virulence factor when present in mammals.

#### **1.2.4.5 Other surface exposed proteins**

OspD is a 28 kDa surface exposed lipoprotein that, like Osps A-C is located on a linear plasmid (Li *et al.*, 2007) OspD appears to be largely downregulated upon the temperature shift associated with the influx of warm mammalian blood when a tick takes a bloodmeal. This suggests that OspD, like OspA and OspB is present on the surface of *Borrelia* when present in the midgut of a tick, and not in mammals. OspD was found to bind the midgut of ticks, and *B. burgdorferi* s.s. deficient in OspD was shown to be less able to colonise ticks, but this didn't have an effect of infectivity on mice (Li *et al.*, 2007).

OspF is another plasmid encoded surface exposed lipoprotein from *Borrelia* (Lam *et al.*, 1994), only it has been shown to not bind Factor H or related proteins (Hovis *et al.*, 2006). OspF is poorly understood and the current function is unknown.

VlsE is a 35 kDa variable surface lipoprotein expressed on the surface of *Borrelia*, the loss of which reduces infectivity (Eicken *et al.*, 2002). VlsE is a predominantly  $\alpha$ -helical protein, consisting largely of a variable cassette region, flanked by the C and N termini of the polypeptide chain. Segments of a silent cassette region can recombine with the cassette region through gene conversion, with nine to thirteen recombination events occurring every four weeks. (Eicken *et al.*, 2002) This is thought to hinder the hosts ability to generate an effective immune response and thus aids the Spirochaete in evasion.

Decorin and fibronectin (Fn) binding proteins (DbpA, DbpB and BBK32) are expressed in early mammalian infection (Guo *et al.*, 1998, Pal *et al.*, 2004a). Fn is a multifunctional dimeric glycoprotein present in blood plasma and on the surface of mammalian cells. Cell

bound Fn aids tissue repair and promotes cell-cell adhesion, (Proctor, 1987, Potts and Campbell, 1996, Mosher, 2001, Kim *et al.*, 2004, To and Midwood, 2011) the binding of which facilitates *Borrelial* transmission into the cell (Steere *et al.*, 2004). Decorin is a glycosaminoglycan present on collagen fibrils (Bunikis *et al.*, 1995, Steere *et al.*, 2004), the binding of which by DbpA & B facilitate the adhesion of the Spirochaete to the mammalian extracellular matrix (Hübner *et al.*, 2001). The Dbps are differentially expressed on the surface of *Borrelia* in response to changes in environmental conditions (Pal *et al.*, 2000, Kim *et al.*, 2004). When *Borrelia* are present in the midgut of tick, neither BBK32 or DbpA or DbpB are expressed, however they are upregulated upon the increase in temperature and pH change that occurs with the influx of mammalian blood during a tick bloodmeal (Neelakanta *et al.*, 2007) and are prominent on the outer when the Spirochaete is present in mammals (Pal *et al.*, 2004b).

A small number of putative porins have been identified to be present in the outer membrane of *Borrelia*; P13 (Skare *et al.*, 1997, Bárcena-Uribarri *et al.*, 2010), Oms28, P66 and a protein homologous to Oms38 in Relapsing fever *Borrelia* (Bárcena-Uribarri *et al.*, 2010, Ristow *et al.*, 2012). P13 and Oms28 are poorly characterised and are only speculated to be porins. P66 is a 66 kDa porin and adhesion present in the outer membrane of *Borrelia* (Skare *et al.*, 1997, Bárcena-Uribarri *et al.*, 2010) that is expressed during mammalian infection (Defoe and Coburn, 2001). P66 is a putative  $\alpha$ -barrel with surface exposed domains, potentially conferring immunogenicity, and is capable of binding  $\alpha$ 3-integrin chains found on activated immune cells, platelets and endothelial cells (Defoe and Coburn, 2001, Ristow *et al.*, 2012), therefore mediating interactions between host cells and *Borrelia*. As was earlier described, P66 is present in lipid rafts present on the surface of *Borrelial* outer membranes (LaRocca *et al.*, 2010).

Enolase is a moonlighting protein; these proteins exist in many varied locations in cells, fulfilling a variety of roles (Jeffery, 1999). *Borrelial* enolase (also known as phosphopyruvate dehydratase (Nogueira *et al.*, 2012)) is surface exposed, known to bind plasminogen, play an important role in enhancing dissemination of *Borrelia* in mammals (Floden *et al.*, 2011) and aid the survival of *Borrelia* in ticks; immunisation with recombinant enolase was shown to reduce Spirochaetal acquisition in feeding ticks (Nogueira *et al.*, 2012). Plasminogen is an inactive single chain glycoprotein present in the hosts serum; and upon binding to *Borrelia* it is cleaved to form plasmin (Toledo *et al.*, 2011); *Borrelia* with plasmin bound on the surface

are capable of degrading Fn and cellular matrices, and penetrating endothelial cells (Floden *et al.*, 2011, Toledo *et al.*, 2011).

### **1.2.5 *Borrelia* implicated with Alzheimer's disease.**

While it is known that *Borrelia* cause Lyme disease and Relapsing fever, Spirochaetes including *Borrelia* and *Treponema* have, along with other pathogens been implicated to be involved in Alzheimer's disease (Miklossy *et al.*, 2005). Although it is not clear whether Spirochaetes are present in the brain before or after the disease is developed there are apparent links between the two. The clinical and pathological characteristics of Alzheimer's disease are comparable to those present in dementia paralytica (the tertiary stage of syphilitic infection) caused by *T. pallidum* which is akin to the manner in which *B. burgdorferi* s.s is capable of causing dementia in late stages of Lyme disease (Miklossy *et al.*, 2005, Swerdlow, 2007, Miklossy, 2008, Miklossy, 2011). Degeneration of the cortex associated with dementia also occurs in the late stages of *B. Burgdorferi* s.s. induced Lyme disease (Miklossy, 1993)

The pathogenic bacteria *Chlamydia pneumonia* and herpes simplex virus type I have also been implicated in infiltrating the central nervous system and causing a neuroinflammatory response, possibly triggering Alzheimer's disease pathology, or bi-polar disorder (Itzhaki *et al.*, 2004, Miklossy, 2008).

Alzheimer's disease is the most common form of dementia in which cognitive and memory functions gradually deteriorate over time. At least part of the cause is suspected to be the accumulation of a 4.2 kDa amyloid beta peptide (A $\beta$ ) in plaque-like structures, and the formation of neurofibrillary tangles in the brain. The cause of this is as yet unidentified (Miklossy, 2011). The Spirochaetes *Borrelia* and *Treponema* have been shown to be present in significant numbers in the brains of deceased Alzheimer's disease sufferers (Itzhaki *et al.*, 2004, Riviere *et al.*, 2002, Miklossy, 2008). It is however, known that OspA present on the outer membrane of *Borrelia* has the capacity to induce amyloid fibres, which are similar to human amyloid (Miklossy *et al.*, 2005, Miklossy, 2011). Several studies have been conducted in an attempt to determine a possible link between *Borrelia* and Alzheimer's disease.

It has been shown that A $\beta$  formed plaque-like structures in several types of mammalian brain cell once exposed to *Borrelia* for up to eight weeks (Miklossy *et al.*, 2005, Miklossy, 2008).

Another study also utilised dark field microscopy show Spirochaetes present in the blood, cerebrospinal fluid and cerebral cortex of 14 Alzheimer's disease patients (and the absence of Spirochaetes in 13 control patients). Peptidoglycan from the *Borrelial* cell walls was co-localised along with the A $\beta$  plaques in 17 Alzheimer's patients, but were again absent in control samples. This, along with the finding of *Borrelial* antigens and genes again co-localised with A $\beta$  plaques provides further evidence for the involvement of *Borrelia* in Alzheimer's disease (Miklossy, 2008).

A study was conducted (Riviere *et al.*, 2002) on the presence of oral *Treponema* and *Borrelia* in the brain and trigeminal nerve in order to investigate the hypotheses that the trigeminal nerve was a route of translocation of *Treponema* to the brain, and whether there was a statistically significant difference in the numbers of Alzheimer's disease/control patients with Spirochaetes present in the brain. Brain sections (frontal lab cortexes and trigeminal ganglia) of donors were tested for seven *Treponema* species and *B. burgdorferi* by the PCR amplification of species-specific 16S rRNA sequences and *Treponema* specific antibodies. PCR found that the brains of fourteen of sixteen patients with Alzheimer's disease were positive for at least one species of *Treponema*, compared to four of eighteen negative controls. Five Alzheimer's patients were found to have *Borrelia*, while all control samples were negative. All five trigeminal ganglia samples also tested positive. Antibodies detected *Treponema* in fifteen of the sixteen patients with Alzheimer's disease. (Riviere *et al.*, 2002).

The geographical distribution of *Borrelia* compared with Alzheimer's disease however, suggests that *Borrelia* cannot be a factor in all cases of Alzheimer's disease. This, along with the presence of *Treponema* in the brains and CNS of patients with Alzheimer's disease may imply that it is not *Borrelia*, but many genera of Spirochaetes that are involved with Alzheimer's disease. The hypothesis that bacteria (Spirochaetes) play a role in the development of Alzheimer's disease is complementary to the hypothesis that genetic defects occur in Alzheimer's disease patients, indeed it is speculated that genetics control susceptibility to bacterial infection (Miklossy, 2008).

### 1.3 Ixodid ticks

*Borrelia* use hard ticks as reservoirs to maintain their natural enzootic cycle and vectors for transmission to mammalian hosts (Steere *et al.*, 2004). Ticks transmit the greatest number of Spirochaetes and viruses of all blood-feeding arthropods (Heinze *et al.*, 2012) and are second only to mosquitoes as vectors of human disease (Parola and Raoult, 2001). The most common *Borrelial* tick reservoirs are species of the Ixodid tick: *I. pacificus* and *I. scapularis* (known as the black-legged or deer tick), which are prevalent mainly in North America, and *I. ricinus*, prevalent mainly in Europe (Figure 1.17) (Steere *et al.*, 2004). In the United States the distribution of reported cases of Lyme disease corresponds closely with that of *I. scapularis* and *I. pacificus* (CDC, 2011). It has also been shown that *I. scapularis* can spread other bacteria and diseases such as *Babesia microti*, causative protozoan of human babesiosis and *Anaplasma phagocytophilum*, causative agent of Human granulocytic anaplasmosis (Stafford, 2001) (Liu *et al.*, 2012, Horowitz *et al.*, 2012). *Ixodes* ticks are also responsible for the spread of viruses such as tick-borne encephalitis and Powassan encephalitis. (Parola and Raoult, 2001). Other species of tick known to spread Lyme disease Spirochaetes are soft ticks of the genera *Argas* and *Ornithodoros* (Vatandoost *et al.*, 2003).



**Figure 1.17. *Ixodes ricinus* suitable habitats in the UK and Europe.** Red shaded areas of the map shows suitable tick habitation. Image adapted from (Gray *et al.*, 2009). Map licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.

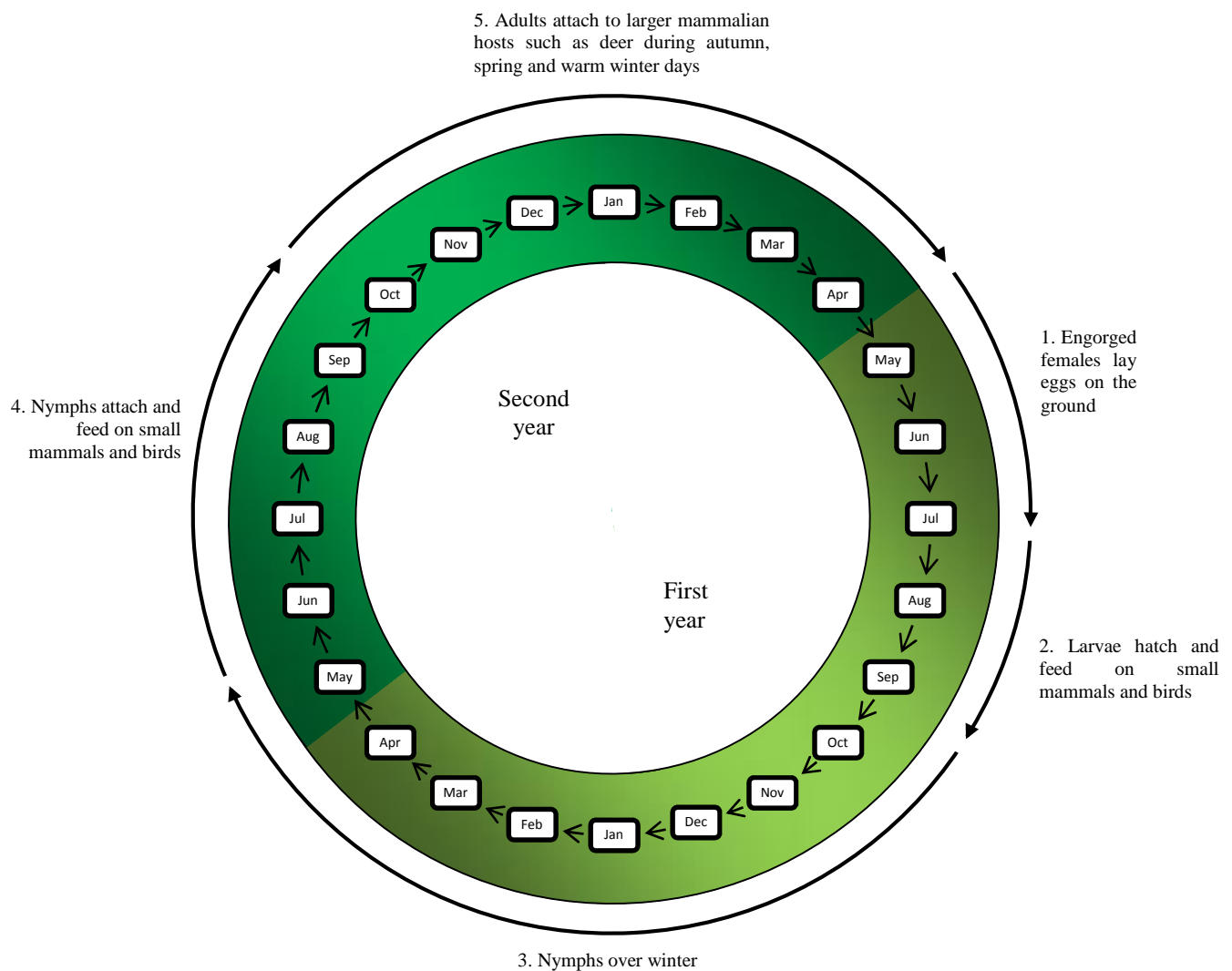
### 1.3.1 Tick life cycle

*Ixodes* ticks have a life cycle which encompasses four stages: egg, larva, nymph and adult (Figure 1.18). The *Ixodid* tick life cycle takes place over the course of approximately two years (Figure 1.19), and begins in the late Spring and early Summer months when females, fresh from a blood-meal lay their eggs. The eggs hatch into larvae over the course of late Summer and the larvae take their first blood-meal, usually on small rodents (in particular the white-footed mouse), lizards or birds before moulting to nymphs over the course of late Autumn through to early Spring (Stafford, 2001). Once Spring has ended, the nymphs take their blood-meal, again on small mammals etc, and moult to adults. These adult ticks feed for the final time on larger mammals such as deer (Schwan and Piesman, 2000) over the course of late Autumn to early Summer, when the engorged females lay their eggs on the ground, thus beginning the cycle again. Whilst some species of tick are known to only feed on specific hosts, *Ixodes scapularis* and *ricinus* feed on a variety of hosts such as small mammals, reptiles and birds when in the first two developmental stages, and large mammals such as deer and dogs (and humans) when adults (Stafford, 2001, Couper *et al.*, 2010, Taragel'ová *et al.*, 2008).



**Figure 1.18. A female and male *I. ricinus*.** Left to Right; Adult female, adult male. The arrow highlight the scutum; this covers the entire dorsal surface in the adult male, however in other developmental stages it is confined to the anterior region of the tick. The adult female is larger than the adult male; this is also true for *Ixodes scapularis* and *pacificus*. Image licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.

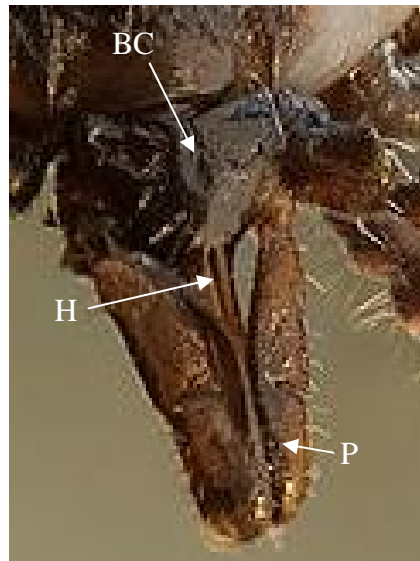




**Figure 1.19.** The two year life cycle of an *Ixodes* tick. Data and images obtained from (Athreya and Rose, 1996, Stafford, 2001)

*Ixodid* ticks have a large (for a tick) body size, varying from 2-30 mm. At the adult and nymphal stages of development, the ticks have four pairs of walking legs, whilst the larval stage has three. Ticks also do not have a distinct head, thorax or abdomen. The mouthparts, along with the hypostome and cutting organs are present on the anterior part of the tick body, also known as the capitulum. The hypostome, labelled H on the scanning electron micrographs (Figure 1.20), has numerous 'barbs' that act as an anchor once the hypostome has penetrated the skin. This makes it difficult to remove the entire tick without the anterior part of the tick body breaking off and remaining lodged in the mammal's skin. Adult ticks can become considerably engorged during a blood meal (Stafford, 2001) with the tick's

weight capable of increasing up to 1000 fold compared to the initial size (Ribeiro *et al.*, 2006).

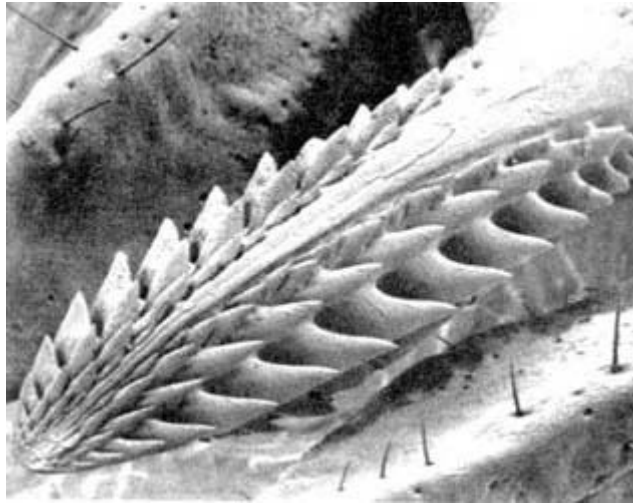


**Figure 1.20. The anterior morphology of *I. ricinus*.** Labelled are: BC - basis capituli, H - external surface of the hypostome, P - palp. Image licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.

*I. scapularis* feeds at three points during its lifetime, once during each of the three stages shown in Figure 1.19. Each blood meal lasts for several days and it is during this time that ticks infected with *Borrelia burgdorferi* s.s. transmit the bacteria to a mammalian host. The most common feedings at which transmission occurs are the nymph and adult stages, as the tick first has to become infected by feeding on a small infected mammal (Stafford, 2001). Once feeding is complete, the tick detaches to complete digestion and either moult to the next stage, or enter a state of reduced metabolism, known as a diapause.

### **1.3.2 The influence of tick saliva on survival of *Borrelia* in hosts**

Many host defences must be overcome by the tick in order to feed, and for successful transmission of the Spirochaete to the mammalian host. The initial obstruction is the mammal's skin; this forms a physical barrier that needs to be penetrated in order to allow access to the bloodstream. The tick achieves this with the hypostome (Figure 1.21).



**Figure 1.21. *Ixodid* tick hypostome.** The ‘barbs’ on the hypostome are present all around the upper half of the hypostome, which are responsible for anchoring the tick into the skin. Image from {Moorhouse, 1981 #554}, licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license.

Once the hypostome has penetrated the skin and has access to the host’s blood supply (a pool of blood formed by laceration of the skin), the tick secretes a cement like substance, to, in conjunction with the hypostome, anchor the tick in the feeding site and prevent detachment (Ribeiro *et al.*, 2006). Once anchored, the tick starts to salivate. Among the first compounds to be released are proteins and enzymes that alleviate host pain and irritation at the binding site (Tyson *et al.*, 2007), the tick also releases a variety of antihaemostatic compounds to maintain a supply of blood to the feeding site (Chmelar *et al.*, 2012). The vasodilators prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) and prostacyclin (PGI<sub>2</sub>) have been shown to be present in the saliva/salivary glands of several species of *Ixodes* tick, and are postulated to increase flow of blood to the tick mouthparts through dilation of host blood vessels, promoting feeding. PGI<sub>2</sub> is also known to inhibit platelet aggregation, as is apyrase, an enzyme present in significant amounts in tick saliva. Apyrase is responsible for the hydrolysis of ATP and ADP (required for platelet aggregation) to AMP and orthophosphate. These two enzymes combined with PGE<sub>2</sub> are effective at disrupting haemostasis (Ribeiro, 1987, Chmelar *et al.*, 2012).

Once anchored into the host’s skin, and with a steady flow of blood, the tick has to negotiate the varied and rapid immune responses at the host’s disposal. This includes, but is not limited to, the various pathways of the complement immune system and several types of adaptive immune cell that reside within the skin. Dendritic cells capture and process antigens, and in turn stimulate B and T lymphocytes and initiate immune responses by secreting cytokines (Banchereau and Steinman, 1998). Mast cells are the primary responders to allergic reactions,

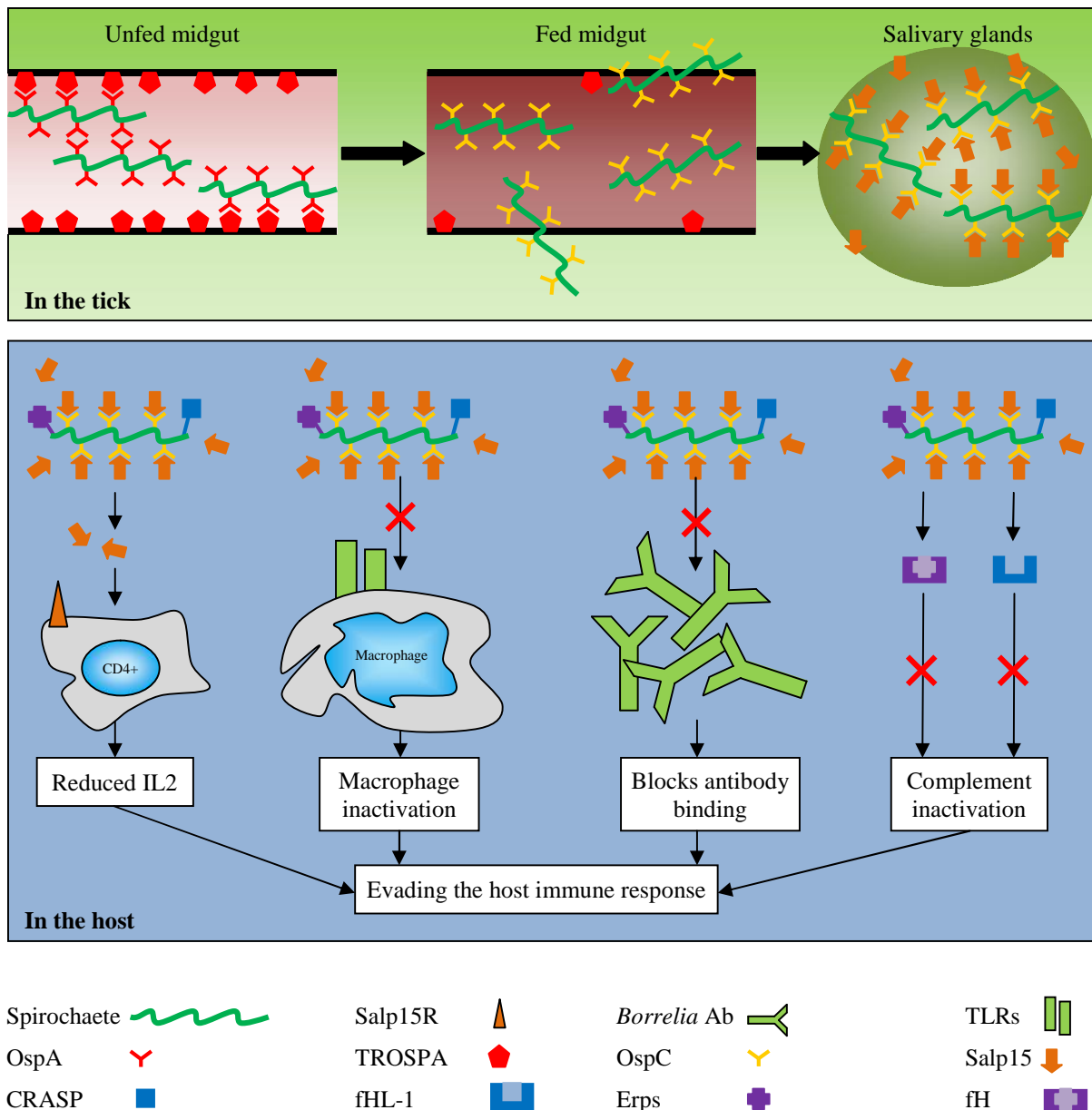
along with being shown to activate inflammatory responses and T-cells (Crivellato *et al.*, 2003, Crivellato *et al.*, 2004). Lymphocytes directly kill microbe infected cells (Brossard and Wikel, 2004). Eosinophils are a source of cytokines and growth factors that display both pro and anti-inflammatory responses such as wound healing (Munitz and Levi-Schaffer, 2004). Macrophages, which are white blood cells responsible for phagocytosis of foreign bodies within the cell (Sarantis and Grinstein, 2012). Keratinocytes compose approximately 95 % of the outer layer of the skin and form the physical barrier to pathogens and non-biological factors such as UV-radiation, and repair of wounds (Nakamizo *et al.*, 2012).

Tick saliva is a well-characterised, complex cocktail of different proteins, enzyme and inhibitors that plays a vital role in circumventing and modulating the host's immune response, facilitating and enhancing not only the tick's ability to feed as described above, but also the conditions for viral infection and dissemination (summarised in Figure 1.22) (Guo *et al.*, 2009, Heinze *et al.*, 2012). Tick saliva contains an arsenal of anti-complement proteins including *Ixodes ricinus* anti-complement proteins (IRAC) I and II (Gillet *et al.*, 2009), Salp15 (Anguita *et al.*, 2002, Juncadella *et al.*, 2008), Salp20 (Tyson *et al.*, 2007), TROSPA (Pal *et al.*, 2004a), vasodilators, immunosuppressive, anti-inflammatory, anti-haemostatic substances and anti-clotting proteins all of which play a role in enabling tick feeding and pathogen transmission (Ribeiro *et al.*, 1985, Ribeiro, 1987, Burkot *et al.*, 1994, Ribeiro *et al.*, 2006, Skallová *et al.*, 2008) *Borrelia* are transmitted from the tick to the host through the tick salivary glands and benefit from tick anti-complement activity. Salivation is not a continuous process throughout the blood meal, taking place every five to fifteen minutes and alternating with ingestion of blood (Ribeiro, 1987). This minimises the formation of antigen-antibody complexes and interactions with complement components in the host blood, which would be deposited at the feeding site, promoting lesions and an immune response (Wikel, 1996). The protein content of saliva varies throughout salivation, allowing the feeding site to be 'primed' prior to the introduction of infectious agents (Heinze *et al.*, 2012); not all pathogens are transmitted at the same time point during the feed. *B. burgdorferi* s.s. is transmitted between 24 and 48 hours after the commencement of the blood meal, whereas other pathogens are transmitted either earlier or later in the blood meal (Ribeiro *et al.*, 2006)

Tick saliva has been shown to: inhibit the maturation and migration of dendritic cells at the binding site, production of cytokines, phagocytosis, T and B cell proliferation and activation of the complement pathway (Skallová *et al.*, 2008, Heinze *et al.*, 2012). Dendritic cell

maturation, differentiation, migration, Th1 and Th2 cytokine production and activation of T-lymphocytes have been shown to be inhibited by proteins in tick saliva, including PGE<sub>2</sub> (Hannier *et al.*, 2003, Skallová *et al.*, 2008, Guo *et al.*, 2009). Proteins in tick saliva have also been shown to reduce proinflammatory responses and Toll-like receptor signalling in macrophages – Salp15 has been shown to bind TLRs. (Anguita *et al.*, 2002, Chen *et al.*, 2012). Mast cells are manipulated by the tick to increase the blood flow to the feeding site; histamine-releasing factor is secreted, triggering mast cells to release histamine and increase vasodilation of surrounding blood vessels (Chmelar *et al.*, 2012). Polymorphonuclear leukocytes have been shown to have inhibited binding of Spirochaetes when pre-treated with tick saliva (Guo *et al.*, 2009).

Tick salivary gland protein Salp15 is a 15 kDa immunomodulatory protein, known to inhibit early events in CD4<sup>+</sup> T-lymphocyte activation (Figure 1.22). It has been shown that Salp15 treated T-cells have reduced calcium signalling within the cell, due to a capability to bind CD4<sup>+</sup> at specific domains (D1 & D2) preventing differentiation into effector T-cells (Paveglio *et al.*, 2007, Juncadella *et al.*, 2008). It is also known to inhibit T-cell and B-cell proliferation, IL-2 production and Natural Killing cell function (Paveglio *et al.*, 2007, Slámová *et al.*, 2011). Salp15 has been shown to bind *B. burgdorferi* OspC and significantly enhance transmission of the Spirochaete during tick feeding, it was also shown that Salp15 secretion is increased when *Borrelia* infected ticks are partaking in a bloodmeal, compared to uninfected ticks (Ramamoorthi *et al.*, 2005, Yang *et al.*, 2005).



**Figure 1.22. Tick protein interactions with *Borrelia* and mammalian host evasion.** In the unfed midgut of the *Ixodes* tick, *Borrelia* expresses OspA, which binds TROSPA and OspA. Once the tick commences a bloodmeal OspA is downregulated and OspC upregulated for transmission to the salivary gland where Salp15 is acquired. Salp15 inhibits the activation of CD4<sup>+</sup> T cells (Anguita *et al.*, 2002). The binding of Salp15 prevents the recognition of antigenic OspC by TLRs and antibodies. The binding of fH and fHL-1 prevents complement activation.

Salp20 inhibits the alternative pathway of the complement immune system by dissociating components of C3 convertase, preventing cleavage of C3 to C3b and therefore deposition of C3b on foreign cell surfaces and therefore lysis, in a similar role to factor H and factor H-like protein 1. This is two-fold in effect, aiding both the tick and any disseminating *Borrelia* by disrupting the complement alternative pathway. It was shown that *B. garinii* incubated with Salp20 had a survival rate of approximately 70 % when faced with alternative pathway

activated complement mediated killing. Salp20 was, however, ineffective against complement activation by the classical pathway. (Tyson *et al.*, 2007).

## 1.4 The Complement System

### 1.4.1 Activation of the complement system

The complement system was discovered in the 1890's as a component of the immune system found to assist antibody mediated killing of bacteria (Sarma and Ward, 2011). The complement system is a factor of both the innate and adaptive immune responses (Fujita, 2002, Schifferli, 2005, Lambris *et al.*, 2008, Sarma and Ward, 2011), capable of swiftly responding to invading pathogens and assisting in long-term immunity. The complement immune response is activated by three distinct pathways; the classical, the alternative and the lectin (Fujita, 2002, Lambris *et al.*, 2008), each of which results in a proteolytic cascade and formation of the membrane attack complexes (MAC) which act to lyse foreign cells (Figure 1.23) (Fujita, 2002). The cascade is a well-controlled network comprised of over thirty proteins that are either present on the cell membranes, or free in plasma (Fujita, 2002, Lambris *et al.*, 2008, Schneider *et al.*, 2009, Sarma and Ward, 2011). In addition to cell lysis by the MAC, activation of the complement system also results in clearance of pathogens by opsonisation (marking pathogens for phagocytosis) and triggering phagocytosis (engulfing and destruction of foreign bodies by phagocytes) (Sarma and Ward, 2011). The complement system also fulfils other functions such as communication with, and activation of adaptive immunity and assisting in the removal of cellular debris (Herbert *et al.*, 2006, Lambris *et al.*, 2008).

The classical pathway is triggered when protein complex C1q recognises antigen-antibody complexes on the surface of a foreign cell (Lambris *et al.*, 2008). The C1q complex comprises one molecule of C1q, two molecules of C1r (a serine protease) and two C1s molecules. Upon binding of C1q to foreign antigens, C1r cleaves C1s, resulting in an activated form that cleaves C2 and C4 (Fujita, 2002).

Activation of the lectin pathway is prompted by recognition and binding of microbial carbohydrate moieties by mannose-binding lectin (MBL) and Ficolin (F). MBL and F are bound to MBL-associated serine proteases (MASPs) present in serum (Fujita, 2002). Some of these MASPs undergo changes in their conformation as a consequence of binding to

pathogens. These changes result in auto activation of MASP2 which is responsible for the cleavage of C4 and C2 (Schifferli, 2005, Sarma and Ward, 2011).

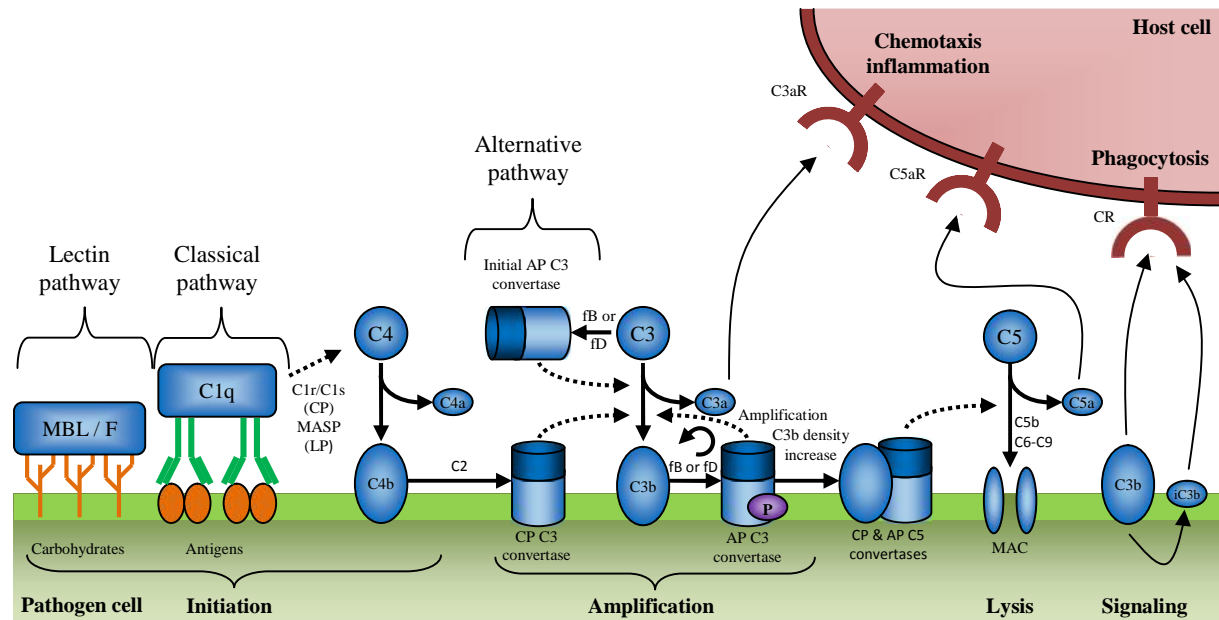
Activation of the complement system by the classical and lectin pathways converge with the degradation of C2 and C4 to produce active fragments C2a, C2b, C4a, and C4b. C4b and C2a combine to form the enzyme complex C4bC2a, known as classical pathway (CP) C3 convertase (Fujita, 2002, Sarma and Ward, 2011). It is at this point that the alternative activation pathway converges with the classical and lectin pathways.

The alternative pathway provides the broadest specificity against pathogens and is not reliant on immunisation or production of antibodies (Pangburn, 2000). Activation of the alternative pathway is selective - host cells and tissues do not trigger activation, whereas foreign organisms do (Sarma and Ward, 2011). The recognition of, and discrimination between host cells and foreign bodies is possible due to a complex set of proteins such as Factor H (fH) and Factor H-like protein (fHL-1). Factor H has multiple sites which can recognise and bind host-like features on the surface of microorganisms and has been shown to have three C3b binding sites. C3 is constantly being hydrolysed at a slow, steady rate, producing small concentrations of C3b, which is capable of binding surfaces lacking FH and fH-like proteins (Sarma and Ward, 2011). The accumulation of C3b on foreign surfaces along with the spontaneous hydrolysis of C3 triggers the alternative pathway of the complement system (Lambris *et al.*, 2008, Sarma and Ward, 2011). As the alternative pathway does not depend on activation by antibodies it is capable of being activated in less than 20 minutes (Lambris *et al.*, 2008).

Each of the complement initiation pathways result in the hydrolysis of C3 to form C3b, which is deposited onto bacterial surfaces (Okemefuna *et al.*, 2008, Sarma and Ward, 2011). Factors B and D bind to the surface C3b, which results in factor D cleaving factor B. One of the products, factor Bb forms a complex with C3b to form C3 convertase, which is stabilised by the binding of properdin, preventing binding and degradation by factors I and H. Meanwhile, the concentration of C3b on the cell surface is increased by continuing hydrolysis of C3. Alternative pathway (AP) C5 convertase is formed by the addition of a further C3b fragment to C3 convertase (C4bC2aC3b). This degrades C5 to produce C5a and C5b. C5b induces the formation of a multi-protein pore-forming membrane attack complex (MAC), incorporating C6 - 9 (C5bC6C7C8C9), which forms a hole in the membrane, lysing the cell (Sarma and Ward, 2011).



C5a, C4a and C3a are anaphylatoxins produced as a result of the activation of the complement cascade. These glycoprotein fragments activate cellular responses such as the release of phagocytes and an inflammatory response (Lambris *et al.*, 2008, Sarma and Ward, 2011).



**Figure 1.23. Activation of the complement immune system by three different mechanisms.** The classical pathway is activated by antibody complexes on the pathogen cell surface, the lectin pathway by recognition and binding of microbial carbohydrate ligands by mannose-binding lectin (MBL) and Ficolins (F) and the alternative pathway by spontaneous and induced hydrolysis of C3. Each of the three activation pathways result in the cleavage of C3 to C3a and C3b. Amplification of C3b amplifies the cascade, mediates immune response by binding complement receptors. C3b build-up and accumulation leads to the assembly of C5 convertase which activate C5a and C5b; the latter of which is responsible for the formation of the membrane attack complex (MAC). Figure adapted from (Lambris *et al.*, 2008)

Each of the complement initiation pathways result in the cleavage of C3 and generation of C3a and C3b. Antibody-antigen complexes recognised by C1q and terminal mannose recognised by MBL trigger the activation and degradation of C4 into C4a and C4b.

Regulation of the complement proteins is vital as accumulation of protein complexes can result in injury to the host tissue (Sarma and Ward, 2011). Complement-mediated lysis of cells is regulated in a multitude of ways, primarily through the rapid inactivation and degradation of components in the presence of cofactors and peptidases and through the prevention of construction and activation of convertases and the MAC. For example, in the presence of factor H, the protease factor I degrades C3b and C4b into several inactive fragments (Okemefuna *et al.*, 2008, Sarma and Ward, 2011). Factor H, in addition to decay acceleration factor (DAF) and C4 binding protein (C4BP) inhibits the assembly and activity



2000, Liang *et al.*, 2004) and any proteins expressed on the surface when present in mammalian hosts have an important function in survival and/or immune evasion (Cassatt *et al.*, 1998, Hübner *et al.*, 2001, Kim *et al.*, 2004, Pal *et al.*, 2004b, Krupka *et al.*, 2011). Surface exposed proteins involved in *Borrelial* immune evasion include: OspC, which binds Salp15, a CD4<sup>+</sup> and lymphocyte inhibitor that also prevents recognition by TLRs and antibodies (Anguita *et al.*, 2002, Juncadella *et al.*, 2008), DbpA, DbpB, enolase and BBK32 which facilitate the binding to, and penetration of mammalian extracellular matrices (Potts and Campbell, 1996, Guo *et al.*, 1998, Brown *et al.*, 1999, Schwarz-Linek *et al.*, 2003).

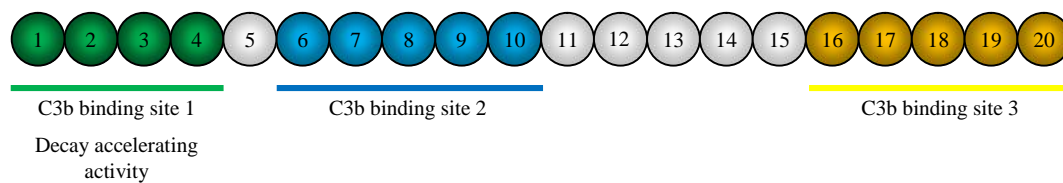
CRASPs (OspE), which bind fH and fH-like proteins are vital to *Borrelia* preventing AP activation of the complement cascade; with fH and fHL-1 bound C3b deposition and accumulation on the cell surface is prevented along with the consequent complement activation (Kraiczy *et al.*, 2006, Rossmann *et al.*, 2006, Bhide *et al.*, 2009, Kraiczy *et al.*, 2009, Sarma and Ward, 2011). Many different bacteria bind Factor H, through a variety of proteins (Shaughnessy *et al.*, 2009) and additional Factor H binding proteins of unknown structure have been identified in *B. afzelii* and *B. garinii* (Bhide *et al.*, 2009). It has been shown *in vitro* that *Borrelia* can switch form from spiral shape to a cystic shape when environmental conditions are unfavourable, switching back to the spiral motile form when conditions are once again favourable; these cystic forms have been found in the brains of deceased Alzheimer's sufferers (Krupka *et al.*, 2011).

*Borrelial* immune evasion is not accomplished entirely by the Spirochaete itself. As was mentioned previously (section 1.3.2), *Borrelial* survival in the mammalian host is aided by several proteins present in tick saliva: Salp20, IRAC, ISAC, which inhibit the formation of C3 convertases and aid in removal of C3b deposits from cell surfaces (Valenzuela *et al.*, 2000, Schroeder *et al.*, 2007, Tyson *et al.*, 2007, Gillet *et al.*, 2009).

#### **1.4.2.1 Human factor H**

Human factor H (fH) is a 155 kDa glycoprotein (Figure 1.25) (Herbert *et al.*, 2006, Okemefuna *et al.*, 2008, Shaughnessy *et al.*, 2009) composed of 20 short consensus repeat (SCR) domains (Kraiczy *et al.*, 2009) responsible for protecting the host cells from self-inflicted damage from the immune system (Coleman *et al.*, 2008). Each SCR domain comprises approximately 60 amino acids, including four conserved disulphide bridge forming

cysteines which are essential for the correct folding of the domain (Prosser *et al.*, 2007, Shaughnessy *et al.*, 2009).



**Figure 1.25. Representation of the domain structure of factor H.** The locations of three C3b binding sites are highlighted (Pangburn, 2000). Factor H is also known to bind heparin in SCRs 7, 9 and 20 modulating attachment to the extracellular matrix. Apoptotic or necrotic cells bind fH through a number of ligands to protect the cells from autoimmunity; Histones for example bind SCRs 1-4, 6-8 and 8-15 (Kopp *et al.*, 2012). It is the four N-terminal SCR domains (1-4) of fH that play a role in the regulating complement responses such as cofactor activity and degradation activity, SCRs 19 and 20 allow attachment of factor H to host cells for inhibition of complement activation at the cell surface (Kopp *et al.*, 2012).

Factor H (along with fHL-1) is the major soluble regulator of the AP of complement activation (Kraiczy *et al.*, 2009, Shaughnessy *et al.*, 2009), and one of the major components of blood plasma, present at a concentration of 200 - 300 µg/ml. The only complement protein present in a higher concentration is C3 (1200 µg/ml). fHL-1 has seven SCR domains and is derived from the fH gene (Kraiczy *et al.*, 2006); these seven domains are identical with the exception of the four C-terminal residues (Pandiripally *et al.*, 2002) fH and fHL-1 degrade and inhibit the formation of C3 convertases, and are cofactors of factor I-mediated degradation of C3b (Bhide *et al.*, 2009, Kraiczy *et al.*, 2009, Shaughnessy *et al.*, 2009). In addition to controlling interactions between C5 and C9 on the cell surface. As was described in the previous section, acquisition of fH is an important feature of bacterial complement evasion and many bacteria have evolved proteins with this function. *Borrelia* are known to bind fH and fHL-1 through CRASPs, Erps and OspEs (Hovis *et al.*, 2006, Kraiczy *et al.*, 2009).

Certain SCRs have been identified as being targeted for pathogenic binding; SCR4 by *S. pyogenes* (Reuter *et al.*, 2010), SCR6 in *N. meningitidis* and *N. gonorrhoeae* (Schneider *et al.*, 2009, Shaughnessy *et al.*, 2009) and SCRs 19-20 in *S. aureus* (Haupt *et al.*, 2008). It is known that *Borrelia* bind fH using BbCRASPs 1 and 2 through SCR 7 (Reuter *et al.*, 2010). One study (Bhide *et al.*, 2009) found that *B. garinii* contained a novel ~19 kDa protein that bound human fH, identified as 'BG0407'. *In silico* analysis conducted by the original authors

evidence for this protein containing two coil-coiled formations, although our research indicates that the protein is a TM  $\beta$ -barrel (chapter 3).

## 1.5 Outer Membrane Proteins

Outer membrane proteins (Omps) are  $\beta$ -Barrel proteins that span the outer membranes of all Gram-negative bacteria, mitochondria and chloroplasts (Schulz, 2000). The number of  $\beta$ -strands varies between eight and twenty two (Khalid *et al.*, 2006). There are multiple types of TM  $\beta$ -barrel OM protein, all of which follow the same basic principles of construction as set out below (Schulz, 2000). Aromatic amino-acids form an 'aromatic girdle' which sits in line with the phosphorous hydrophilic outer layers of the membrane and 'protects' the inner core of hydrophobic amino-acids in the centre of the strands (Arora *et al.*, 2000).

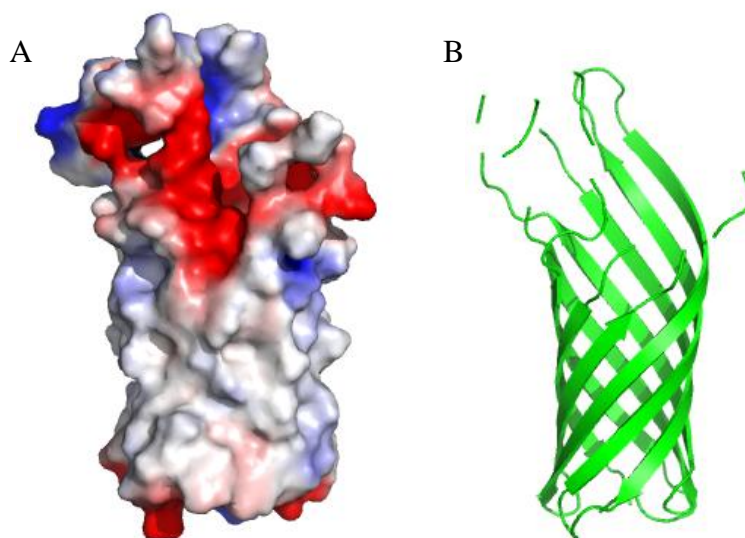
- Both the N and C termini are located at the periplasmic end of the barrel
- The number of strands in a  $\beta$ -barrel is even
- The  $\beta$ -strands are tilted at an angle of approximately 45 °; only one direction is assumed in each protein as the other is not energetically favoured
- The shear number of a  $\beta$ -barrel is the generally  $n + 2$ , where  $n$  = the number of strands
- All strands are connected to neighbouring chains and are anti-parallel to each other
- The connections between the strands at the periplasmic end are designated as turns and numbered T1, T2 etc; each turn is usually only a couple of residues long
- The connections between the strands at the external barrel end are usually longer loops and are designated L1, L2 etc
- The surface of the  $\beta$ -barrel in contact with the non-polar interior of the membrane mostly consists of residues with non-polar side chains
- Around this non-polar section of the barrel, are hydrophilic side chains that contact the area of the membrane where the non-polar interior meets the polar exterior
- $\beta$ -barrel proteins have high evolutionary sequence variability
- The external loop regions of the protein are mobile and have very high sequence variability

In addition to the principles laid out above, for  $\beta$ -barrel formation at least four of the five residues with side chains pointing towards the membrane must be hydrophobic, with none of the three central residues charged and residues on the inside of the barrel must have small

side chains (Koebnik, 1999). OM proteins fulfil many roles in cell membranes and cellular interactions of a wide range of Gram-negative bacteria, such as aiding in structural integrity and adhesion to the extracellular matrices of host cells and (Beher *et al.*, 1980, Prasadarao *et al.*, 1996).

### 1.5.1 OmpA & its role in infection & immune evasion

*E. coli* OmpA is a 35 kDa protein highly conserved protein within Gram negative bacteria (Prasadarao *et al.*, 1996) that comprises a 171 residue (~20 kDa) N-terminal  $\beta$ -barrel membrane-spanning domain (Figure 1.26) (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999) and a globular C-terminal periplasmic domain (Power *et al.*, 2006) composed of approximately 150 residues (Arora *et al.*, 2000). OmpA is known to be involved in structural stability, and translocation and invasion of endothelial cells (Prasadarao *et al.*, 1996, Kim, 2001). OmpA expression is tightly regulated at the posttranscriptional level (Smith *et al.*, 2007). The entire protein sequence is 325 amino-acids long, with a 21 residue signal sequence (Freudl *et al.*, 1986).



**Figure 1.26. The TM spanning domain of *E. coli* OmpA.** A) Electrostatic surface of OmpA. B) Cartoon structure of OmpA. Structure solved by X-ray crystallography at a resolution of 2.50 Å (Pautsch and Schulz, 1998), RCSB PDB accession 1BXW. Images produced using PyMol (Schrödinger, 2010).

OmpA from *E. coli* has been described as a heat-modifiable protein, due to the change in migration on an SDS-PAGE. Whole OmpA pre-boiled above 60 °C in the presence of SDS has been reported to migrate at a molecular weight of 35- 36 kDa (Beher *et al.*, 1980, Freudl *et al.*, 1986), whereas OmpA not boiled migrates at 28 – 31 kDa (Freudl *et al.*, 1986). This is

thought to be due to a conformational change (denaturing) when boiled and the native protein binding excess amounts of SDS (Freudl *et al.*, 1986). *E. coli* OmpA is a known bacteriophage (Fernández *et al.*, 2001) and colicin receptor (Beher *et al.*, 1980, Prasadarao *et al.*, 1996) through the outer loops (Buchanan, 1999).

The structure of the N-terminal  $\beta$ -barrel domain has been solved by both X-ray crystallography (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Pautsch and Schulz, 2000) and NMR spectroscopy (Arora *et al.*, 2001), with the structures agreeing to a high degree. Minor differences are found in the length of the  $\beta$ -sheets of OmpA in solution compared to crystals, but the overall barrel shape remains the same (Arora *et al.*, 2001). Previous analysis of the peptide sequences has found that the  $\beta$ -strands are well conserved, whereas the loop regions are highly variable (Power *et al.*, 2006). The C-terminal domain of OmpA is composed of residues 172-325 and is known to interact with periplasmic peptidoglycan (Buchanan, 1999, Power *et al.*, 2006), helping maintain the structural integrity of the cell (Hong *et al.*, 2006).

It has been shown that OmpA has multiple functions in *E. coli*: in addition to aiding structural integrity and maintaining shape (Prasadarao *et al.*, 1996, Jeannin *et al.*, 2002), it plays an important role in mediating pathogenesis and virulence of bacteria in humans (Fernández *et al.*, 2001). This includes adherence to, and invasion of mammalian cells (Fernández *et al.*, 2001, Jeannin *et al.*, 2002) and facilitating translocation of the blood-brain barrier (BBB) through brain microvascular endothelial cells (BMECs) (Kim, 2001). Translocation of the BBB has been shown in neonatal meningitis-causing *E. coli* (Prasadarao *et al.*, 1996). *E. coli* OmpA is known to bind the 96 kDa glycoprotein Ecgp present on the surface of mammalian cells (Niemann *et al.*, 2004), which is a potential as a trigger for the accumulation of cellular actin under the host cell surface, and the endocytic process by which translocation occurs (Prasadarao *et al.*, 1999). OmpA in *Cronobacter* has been shown to aid invasion of human intestinal epithelial cells (Kim *et al.*, 2010). The membrane spanning domain of *E. coli* OmpA has sequence homology to one protein expressed on the surface of *N. gonorrhoeae* thought to be involved in epithelial cell invasion, while the periplasmic C-terminal domain shares sequence homology to a complement resistance protein also in *N. gonorrhoeae* (Prasadarao *et al.*, 1996) suggesting a similar role in *E. coli* pathogenesis. OmpA has been shown to exhibit large ion channel conductance in vitro, potentially due to the presence of hydrophobic residues lining the centre of the barrel (Hong *et al.*, 2006). This is in contrast to

the crystal structure, which shows no continuous channels consistent with a transport function. A possible explanation for this would be a molecular ‘gate’, initially formed by strong electrostatic interactions between residues Arg138 and Glu52 blocking ion transport. This ‘gate’ is speculated to be ‘opened’ by formation of electrostatic interactions between Glu52 and Lys82 instead of Arg138 (Smith *et al.*, 2007). *V. cholera* have been shown to control the release of OM vesicles by limiting expression of OmpA as a mechanism of defence against UV damage; repression of OmpA expression results in an increased number of vesicles being produced, providing further evidence for the role of OmpA stabilising the OM (Song and Wai, 2009). Small TM  $\alpha$ -stranded proteins have also been implicated with lipid metabolism in the OM of bacteria (Hong *et al.*, 2006).

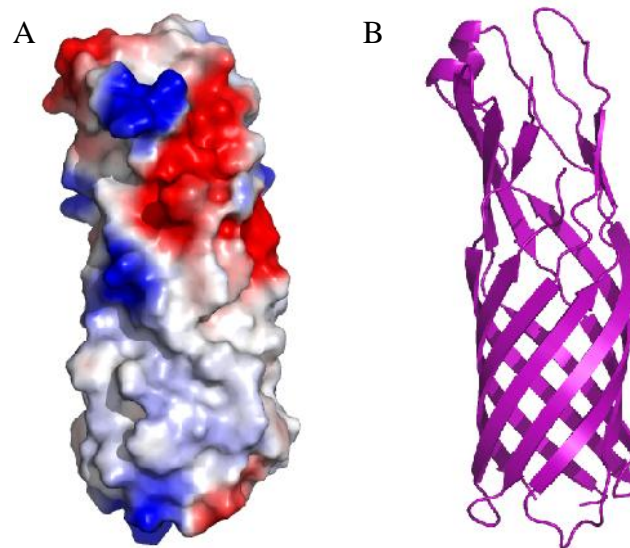
*E. coli*, *N. meningitidis*, *H. influenza* type b and *S. pneumoniae* meningitis causing pathogens are able to pass from the blood of the infected host to the central nervous system (CNS) without affecting the structure and reliability of the BBB (Kim, 2001). It has been shown that the N-terminal loop regions of OmpA that protrude from the surface of the bacterium contribute to invasion of brain micro vascular endothelial cells (BMECs) and that OmpA positive (OmpA<sup>+</sup>) strains of *E. coli* were 25-50 fold more efficient than OmpA<sup>-</sup> strains at invading BMECs in vitro (Prasadarao *et al.*, 1996). It was also found that BMECs incubated with purified OmpA, synthetic OmpA loop regions and OmpA antibodies prior to the introduction of OmpA<sup>+</sup> *E. coli* caused significant inhibition of the invasion of OmpA<sup>+</sup> strains of *E. coli* (Prasadarao *et al.*, 1996). OmpA from *E. coli* have been shown to induce humoral and cytotoxic responses, and it was found that OmpA deficient *E. coli* have reduced virulence and capacity to invade mammalian cells (Jeannin *et al.*, 2002). OmpA from *E. coli* has been shown to bind complement binding protein 4 (C4bp), which is an inhibitor of C3b activation via the classical complement activation pathway and to aid in biofilm formation (Smith *et al.*, 2007).

### 1.5.2 OmpW

OmpW is a 192 residue, 21 kDa  $\alpha$ -barrel protein that spans the outer membrane of *E. coli* (Figure 1.27) (Albrecht *et al.*, 2006). OmpW has been shown to display the properties of a hydrophobic channel (Hong *et al.*, 2006). The centre of the barrel of OmpW is largely hydrophobic and has a large hydrophobic binding pocket in the centre of the extracellular loop regions which extends towards the centre of the barrel. This ends at a ‘gate’ comprised



of a leucine and a tryptophan residue approximately halfway in the membrane spanning region, just below a gap in the outer surface of the barrel postulated to be the channel exit (Hong *et al.*, 2006). Hydrophobic molecules generally have difficulty passing through the LPS layer on most bacterial surfaces, hence the need for a hydrophobic transporter, potentially delivering molecules to the middle of the membrane via the hydrophobic channel and a gap in the side of the barrel (Hong *et al.*, 2006).



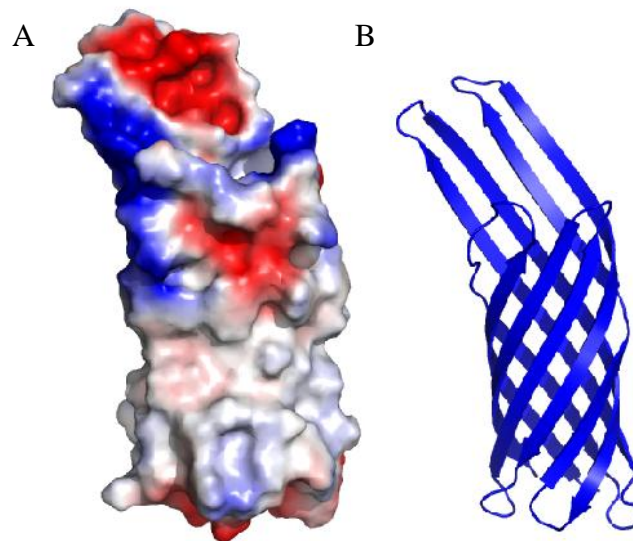
**Figure 1.27. *E. coli* OmpW.** A) Electrostatic surface of OmpW. B) Cartoon structure of OmpW. The structure was solved in 2005 at a resolution of 3.0 Å (Hong *et al.*, 2006). Data retrieved from the RCSB PDB using accession 2F1T, images produced using PyMOL (Schrödinger, 2010).

OmpW possesses a binding site similar to glycerol binding sites in other proteins (Hong *et al.*, 2006). This is unusual as the centre of  $\beta$ -barrels are generally hydrophilic. A small number of proteins are known to have a hydrophobic binding pocket and these bind hydrophobic molecules (Hong *et al.*, 2006). OmpW has been shown in *E. coli* to be regulated by the availability of iron, and upregulated in *E. coli* infection of mice (Albrecht *et al.*, 2006); homologues of OmpW are thought to be involved in adaptation to stress conditions (Hong *et al.*, 2006). OmpW from *V. cholerae* has been shown to exhibit immunogenic properties (Albrecht *et al.*, 2006).

### 1.5.3 OmpX

OmpX is an 18 kDa, 172 residue membrane spanning  $\beta$ -barrel protein with a 23 residue N-terminal signal sequence (Figure 1.28) (Stoorvogel *et al.*, 1991, Vogt and Schulz, 1999). OmpX has eight internal cavities within the barrel (though none of these span the barrel),

composing a H-bonding network that incorporates several charged residues; the barrel of OmpX resembles an inverse micelle (Vogt and Schulz, 1999).



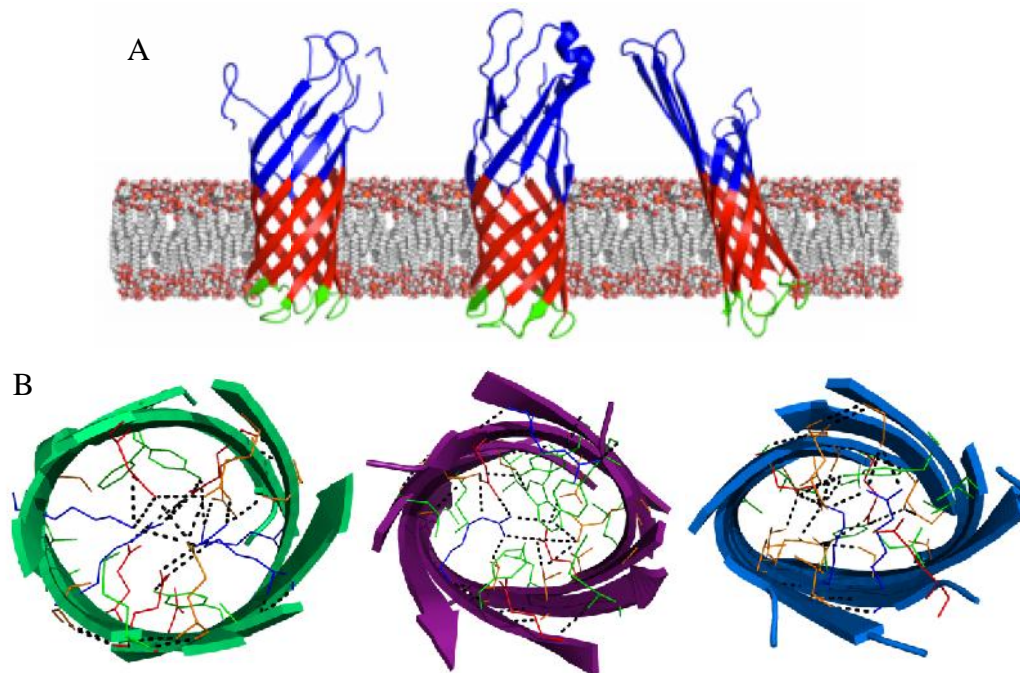
**Figure 1.28. *E. coli* OmpX.** A) Electrostatic surface of OmpX. B) Cartoon structure of OmpX. Structure solved by X-ray crystallography at a resolution of 1.9 Å in 1999 (Vogt and Schulz, 1999), data retrieved from the RCSB PDB (accession 1QJ8). Images produced using PyMOL (Schrödinger, 2010).

OmpX is a mediator of mammalian cell adhesion, a regulator of complement evasion (Fernández *et al.*, 2001) and is involved in neutralising host defence mechanisms (Kim *et al.*, 2010). *E. coli* OmpX is homologous to membrane proteins PagC and Rck (from *Salmonella typhimurium*), Ail (from *Yersinia enterocolitica*) and the bacteriophage Lom; these proteins enable the respective bacteria to adhere to mammalian cells, and inhibit lysis by complement (by preventing formation of the MAC) (Pautsch *et al.*, 1999, Vogt and Schulz, 1999). Studies have shown that an increase in OmpX expression in *E. cloacae* increases the infectiveness of this bacteria in rabbits (Vogt and Schulz, 1999). The  $\beta$ -sheets that protrude into the extracellular space are rigid unlike those in OmpA and may be used to facilitate mammalian cell adhesion and complement evasion by binding to proteins with a complementary  $\beta$ -strand (Vogt and Schulz, 1999, Schulz, 2000).

#### 1.5.4 A structural comparison of OmpA, OmpX and OmpW

OmpX, OmpW and the transmembrane domain of OmpA each consist of eight anti-parallel  $\beta$ -sheets (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Pautsch and Schulz, 2000, Albrecht *et al.*, 2006, Hong *et al.*, 2006), each with four extracellular loops and three periplasmic turns (Figure 1.29 A). OmpW has a small  $\alpha$ -helical region present in L3 (Hong *et al.*, 2006) which

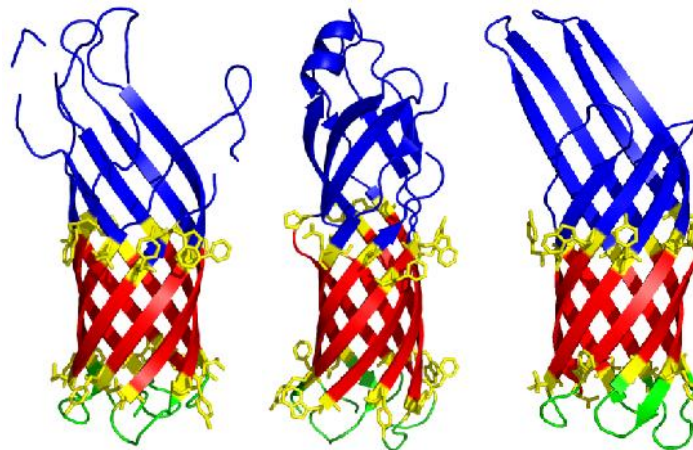
the OmpA and OmpX do not. OmpX shares the same  $\beta$ -barrel topology as OmpA and OmpW, but has a different shear number (8 compared to OmpA and OmpW, which both have a shear number of 10) and different hydrogen bonding within the centre of the barrel (Vogt and Schulz, 1999, Pautsch and Schulz, 2000, Hong *et al.*, 2006). Each of the proteins has an internal H-bonding network (Figure 1.29 B) that appears to leave very little space for molecules to pass through the barrel, although as mentioned it is possible that both OmpA and OmpW have an internal molecular gate' (sections 1.5.1 & 1.5.2)



**Figure 1.29. OmpA, W and X from *E. coli*.** **A)** Blue regions are extracellular loop regions and sections of the  $\beta$ -barrel protruding into extracellular space,  $\beta$ -sheets highlighted red are membrane spanning, green regions are periplasmic, extending below the outer membrane. The periplasmic C-terminal domain of OmpA is not included. **B)** Top down view of OmpA, OmpW and OmpX showing the overall barrel shape in addition to hydrogen bonds across the centre. Residues are coloured as follows: red – negatively charged, blue – positively charged, green - hydrophobic, orange – polar uncharged. All protein structure images prepared using PyMOL (Schrödinger, 2010) and .pdb files from the RCSB protein data bank; accession numbers 1BXW, 2F1T and 1QJ8 respectively (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Vogt and Schulz, 1999, Hong *et al.*, 2006). The top-down view suggests that these proteins do not function as channels due to the packing of side-chains within the centre of the barrel.

OmpX, like OmpA and OmpW provides evidence that the membrane-spanning region of the protein is much more highly conserved than the loop regions (Vogt and Schulz, 1999). OmpW has longer extracellular loop regions than OmpA (which in turn has longer loops than OmpX (Fernández *et al.*, 2001)), in addition to one extracellular  $\alpha$ -helix and multiple  $\beta$ -sheets. The periplasmic turn T3 is a very tight turn, and is common in both OmpA and OmpX (Vogt and Schulz, 1999). The final three C-terminal residues of the barrels are tyrosine, any

residue and phenylalanine, this appears to be common to all Gram-negative  $\beta$ -barrel OM proteins and may be essential for folding (Vogt and Schulz, 1999) (section 3.2, Figure 3.6). Another difference is in the top-down shape of the barrel, OmpA is almost circular, OmpW is elliptical and OmpX more elliptical and thinner than OmpW. As mentioned previously, a key feature of the  $\beta$ -barrel is the 'girdle' of aromatic residues (highlighted yellow, Figure 1.30) that orients the protein in the membrane and borders the hydrophobic regions of the  $\beta$ -strands (Vogt and Schulz, 1999, Pautsch and Schulz, 2000, Schulz, 2000).



**Figure 1.30. The aromatic girdle of OmpA, OmpW and OmpX.** L-R: OmpA, OmpW, OmpX. The loops, transmembrane regions and periplasmic turns are coloured blue, red and green respectively. The residues making up the aromatic girdle have been shown as 'sticks' and coloured yellow. All protein structure images prepared using PyMOL (Schrödinger, 2010) and .pdb files from the RCSB protein data bank; accession numbers 1BXW, 2F1T and 1QJ8 respectively (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Vogt and Schulz, 1999, Hong *et al.*, 2006).

## 1.6 Aims of the research

More target proteins expressed by *Borrelia* are required for a number of reasons; the exposure of very few antigenic lipoproteins during mammalian infection means that current diagnostic systems are very unreliable and treatment of middle and late stage Lyme disease / neuroborreliosis is challenging. Identification of more surface exposed proteins during mammalian infection would aid in production of more accurate diagnostic tests, treatments and vaccines, in addition to developing a better understanding of understanding of the molecular mechanisms of dissemination, invasion and immune evasion.

It has been shown that *Borrelia* contain a small number of membrane spanning  $\beta$ -barrels such as P66. This coupled with the existence of an orthologue of the  $\beta$ -barrel assembly apparatus BamA and the Sec pathway, suggests that it is possible that *Borrelia* produce other, so far unknown, integral membrane protein analogous to Gram-negative bacteria. If this is the case,

it is likely that some of these proteins might closely resemble the prototypical OmpA-like membrane spanning domain described previously (section 1.5.1) due to its simplicity and high conservation among other proteobacteria. The fact that this protein has not been identified to date is likely due to the difficulties in predicting  $\beta$ -barrels, which are much harder to identify than lipoproteins, and the fact that *Borrelia* as Spirochaetes are not closely related to Gram-negative bacteria. It also hints at low levels of expression.

OmpA is known to have multiple important cellular functions and be expressed in numerous Gram-negative bacteria throughout mammalian infection. The ability to evade the complement immune system is common to all Gram-negative bacteria and is known to be connected with the OmpA-family of proteins (Vogt and Schulz, 1999). Spirochaetes have a double membrane structure like *E. coli*, although very few TM spanning proteins have been identified to date. It is known that *Borrelia* has a BamA orthologue (Lenhart and Atkins, 2009), used in *E. coli* to assemble the  $\beta$ -barrel at the OM. The results from study (Prasadarao *et al.*, 1996) clearly highlight the important role that outer membrane proteins, and in particular OmpA plays in the invasion of the CNS in *E. coli*. As it has been shown that *Borrelia* also translocate across the BBB to invade the CNS (Athreya and Rose, 1996, Kurtenbach *et al.*, 2006) it is highly possible that OmpA-like proteins are present in *Borrelia* and that they play a similar role. As OmpA from *E. coli* interacts with the peptidoglycan layer and *Borrelia* is structured differently it is more likely that *Borrelia* contain OmpX or OmpW than OmpA, hence our protein is named OmpB. As OmpW functions as a transporter of hydrophobic molecules across the outer membrane it can be reasoned that *Borrelia* should contain OmpW.

In summary the aims of this research were:

- To conduct bioinformatic analysis of the *Borrelial* genome and identify any potential OmpA-like genes.
- Characterise and determine if the protein coded for resembled an OmpA-like protein
- Identify any binding (such as fH or the extracellular matrix)
- Identify any roles with the bacteria.
- Characterise the structure of the protein through various methods such as CD, SAXS and X-Ray crystallography
- Determine if the protein(s) were most similar to OmpA, OmpX or OmpW.

## **Chapter 2. Materials & Methods**

### **2.1 Materials**

DNA samples of *B. garinii* (71186B) and *B. afzelii* (70250B) were supplied by Dr Gabi Margos (The University of Bath) in the form of whole DNA extracted from infected *I. ricinus*. Samples of DNA from *B. burgdorferi* s.s. PKa2, *B. afzelii* PKo and *B. garinii* PRef were supplied by Dr. Volker Fingerle and Cecilia Hizo-Teufel, National Reference Centre for *Borrelia*, LGL, Germany. Cell stocks were purchased from New England Biolabs (Hitchin, UK) and Merck (Merck Serono Ltd, Middlesex, UK). Plasmids were purchased from Novagen (Merck Serono Ltd, Middlesex, UK). All standard lab reagents were purchased from Fisher Scientific UK (Loughborough, UK) or Sigma-Aldrich Company UK (Gillingham, UK).

## 2.1.1 Reagents, buffers and purchased media

Chemical or Reagent	Abbreviation	Supplied by
Acrylamide: Bis-Acrylamide 37.5 : 1 (40 % v/v solution/electrophoresis)	Bis-Tris	Fisher
Agar	-	Fisher
Agarose (DNase and RNase free)	-	Fisher
Ampicillin sodium salt	Amp	Fisher
Ammonium persulphate	APS	Sigma
L (+) arginine hydrochloride	Arg	Fisher
-mercaptoethanol	ME	Fisher
Bovine Serum Albumin	BSA	Sigma
Bradford Reagent	-	Invitrogen
5-Bromo-4-chloro-3-indoyl- $\beta$ -D-galactopyranoside	X-Gal	Fisher
3-[(3-Chloramidopropyl)dimethylammonio]-1-propanesulphonate	CHAPS	Fisher
Chloramphenicol	Chl	Fisher
L(+) cysteine	-	Fisher
Dimethyl sulphoxide	DMSO	Fisher
Disodium hydrogen orthophosphate	-	Fisher
Dithiothreitol	DTT	Fisher
DNase	-	Sigma
Ethanolamine	-	Fisher
Ethidium Bromide	Et-Br	Fisher
Ethylenediaminetetraacetic acid	EDTA	Fisher
Guanidine hydrochloride	GuHCl	Fisher
HEPES crystals	-	Fisher
Imidazole	-	Fisher
Isopropyl $\beta$ -D-1-thiogalactopyranoside	IPTG	Fisher
Kanamycin sulphate	Kan	Fisher
Lauryldimethylamine-oxide	LDAO	Sigma
Lysozyme	-	Sigma
Monoolien	MO	Sigma
n-octyl- $\beta$ -D-glucoside	OG	Sigma
n-octyltetraoxyethylene	C8E4	Sigma
Polyethylene glycol 8000	PEG 8000	Fisher
Potassium phosphate monobasic	-	Fisher
Protease inhibitor cocktail (EDTA free)	-	Sigma
Sodium dodecyl sulphate	SDS	Sigma
Sodium phosphate monobasic	-	Fisher
SYBR-Safe	-	Intvitrogen
Tetramethylethylenediamine	TEMED	Sigma
Tris(hydroxymethyl)aminomethane-base	Tris	Fisher
Tris(hydroxymethyl)aminomethane-HCl	Tris-HCl	Fisher
Triton X-100	-	Fisher
Tryptone	-	Fisher
TWEEN-20	-	Sigma
Urea	-	Fisher
Yeast extract	-	Fisher

**Table 2.1. Chemicals and reagents used in this research.** Standard chemicals not listed were purchased from either Fisher Scientific or Sigma-Aldrich.



Buffer / Media	Composition
50x 5052	25 % (w/v) -lactose, 2.5 % (w/v) glucose, 10 % (w/v) glycerol
Agar plates	LB with 1.5 % (w/v) agar
Agar-Kan plates	LB-Kan with 1.5 % (w/v) agar
Agar-Kan/Chl plates	LB-Kan/Chl with 1.5 % (w/v) agar
Agar-Amp/X-gal plates	LB-Amp with 1.5 % (w/v) agar and 40 µg/ml X-gal
Agar-Amp/Chl/X-gal plates	LB-Amp/Chl with 1.5 % (w/v) agar and 40 µg/ml X-gal
Autoinduction media	1 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 1x NPS, 1x 5052
BSK-H medium, complete	-
Coomassie Blue stain	0.1 % w/v Coomassie R250, 10 % v/v glacial acetic acid, 40 % methanol
Coomassie Destain	10 % (v/v) glacial acetic acid, 20 % (v/v) methanol
Elution buffer	0.5 M NaCl, 0.5 M imidazole, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
FPLC Buffer A	6 M GuHCl, 0.5 M NaCl, 50 mM Tris, pH 8.0
FPLC Buffer B	6 M GuHCl, 0.5 M NaCl, 50 mM Tris, 0.5 M imidazole, pH 8.0
Lysis buffer	500 mM NaCl, 50 mM Tris, 0.4 µg/ml lysozyme, 10 mM -mercaptoethanol, 0.1 % (v/v) Triton X-100, pH 8.0
Lysogeny Broth (LB)	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl
Lysogeny Broth-Amp (LB-Amp)	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 100 µg/ml Amp
Lysogeny Broth-Amp (LB-Amp/Chl)	LB-Amp with 34µg/ml Chl
Lysogeny Broth-Kan (LB-Kan)	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, 50 µg/ml Kan
Lysogeny Broth-Kan (LB-Kan/Chl)	LB-Kan with 34µg/ml Chl
MES running buffer	-
20x NPS	1 M Na <sub>2</sub> HPO <sub>4</sub> , 1 M KH <sub>2</sub> PO <sub>4</sub> , 0.5 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
PBS (1x)	137 mM NaCl, 2.7 mM KCl, 4.3 mM Na <sub>2</sub> HPO <sub>4</sub> , 1.47 mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.4
PBS-TWEEN (0.05%)	1x PBS, 0.05% (v/v) TWEEN-20, pH 7.4
Refolding buffer 1 (RF1)	4 M GuHCl, 0.5 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
Refolding buffer 2 (RF2)	2 M GuHCl, 0.5 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
Refolding buffer 3 (RF3)	1 M GuHCl, 0.5 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
Refolding buffer 4 (RF4)	0.5 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
Resolubilisation buffer	8 M Urea, 0.1 M Tris, 1 mM EDTA, 2 mM DTT, pH 8.0
Transfer buffer	1.45% (w/v) Tris, 7.2 % (w/v) glycine, 20 % (v/v) methanol
Tris/borate/EDTA (TBE)	Supplied as 10x and diluted to 0.5x in distilled water
TBS (1x)	0.15 M NaCl, 20 mM Tris-HCl, pH 7.4
TBS-TWEEN	1x TBS, 0.05 % (v/v) TWEEN-20, pH 7.4
TSS Buffer	10 % (w/v) PEG 8000, 30 µM MgCl <sub>2</sub> .6H <sub>2</sub> O, 1.25 % (v/v) DMSO, LB up to volume
SAXS 0.1 % LDAO buffer	0.3 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
SAXS 0.01 % LDAO buffer	30 mM NaCl, 5 mM Tris, 0.01 % (w/v) LDAO, pH 8.0
Sample Loading Solution (SLS)	-
Size exclusion buffer	0.5 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO, pH 8.0
Source Outgrowth Control (SOC)	-
Media	
Salt buffer	0.3 M NaCl, 50 mM Tris-HCl, pH 8
Wash buffer	6 M GuHCl, 0.3 M NaCl, 50 mM Tris-HCl, 20 mM imidazole, pH 8

**Table 2.2. Buffers and Media used throughout this research.** BSK-H media was purchased from Sigma-Aldrich Company Ltd, SLS from Invitrogen and SOC from NEB.



PCR, restriction digest and ligation reagents	Supplied by
BamHI (20,000 U/ml)	NEB
BenchTop 1000 bp DNA Ladder	Promega
BglII (10,000 U/ml)	NEB
6X Blue/Orange Loading Dye	Promega
DNA Step Ladder (50 bp increments)	Promega
dNTP mix (10 mM each)	Promega
EcoRI (20,000 U/ml)	NEB
GoTaq Flexi Green Buffer (5 x)	Promega
GoTaq Hot Start Polymerase (5 U/μl)	Promega
HindIII (20,000 U/ml)	NEB
MgCl <sub>2</sub> (25 mM)	Promega
NEBuffer 2	NEB
NEBuffer 3	NEB
NEBuffer EcoRI	NEB
T4 DNA Ligase (400,000 U/ml)	NEB

**Table 2.3. PCR reagents, enzymes and buffers used in this work.** T4 DNA Ligase was diluted in nuclease free water to a concentration of 1 U/μl before use. All enzymes were used neat.

## 2.1.2 Primers

All of the primers used in this work were designed using OligoExplorer 1.2 and OligoAnalyzer 1.2 (Genelink) (Table 2.4, Table 2.5 & Table 2.6). Primers were purchased from Eurofins MWG Operon (Ebersberg, Germany). Primers were supplied lyophilised and were diluted to a concentration of 100 pmol/μl using nuclease free water.

Target	Primer	Sequence (5'-3')	Length (bp)	T <sub>m</sub> (°C)
PepX	PepXIF	TTA TTC CAA ACC TTG CAA TCC	21	55.1
	PepXIR	TGT GCC TGA AGG AAC ATT TG	20	56.1
	PepXOF	ACA GAG ACT TAA GCT TAG CAG	21	55.9
	PepXOR	GTT CCA ATG TCA ATA GTT TC	20	51.2
UvrA	UvrAIF	GCT TAA ATT TTT AAT TGA TGT TGG	24	52.5
	UvrAIR	CCT ATT GGT TTT TGA TTT ATT TG	23	51.7
	UvrAOF	GAA ATT TTA AAG GAA ATT AAA AGT AG	26	52.2
	UvrAOR	CAA GGA ACA AAA ACA TCT GG	20	53.2

**Table 2.4. Primers for the nPCR amplification of housekeeping genes PepX and UvrA.** The primers amplify housekeeping genes as positive controls. The sequences supplied by Dr. G. Margos (University of Bath). The first four letters of the primer designation refer to the gene being amplified, and the last two letters of detail that primers function: O is a primer used in the first (outer) amplification of the target gene and a selection of the surrounding sequence, whereas I is used in the second (inner) amplification of the target gene alone. The letters F and R respectively denote whether the primer is forward or reverse.

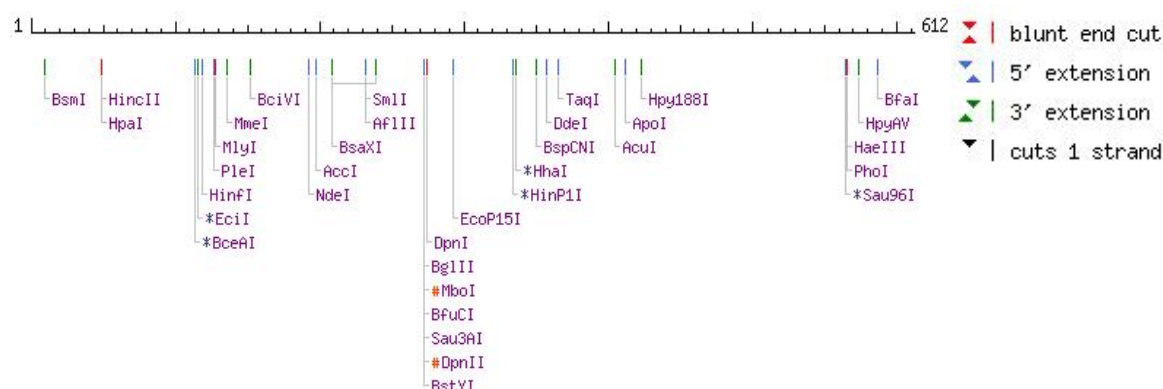
*Borrelial* UvrA is a homologue of a DNA repair protein found in *E coli* and many other bacteria; that is in conjunction with several other proteins responsible for repairing DNA damage through excision of nucleotides (Sambir *et al.*, 2011).

PepX is a dipeptidyl aminopeptidase commonly used along with UvrA and several other proteins as a housekeeping gene in evolutionary genetic studies of *Borrelia* (Margos *et al.*, 2008).

Target	Primer	Sequence (5'-3')	Length (bp)	T <sub>m</sub> (°C)
BAPKO_0026 (Q0SPD7)	A0026IF1	CAA GGG TAA TAA TTT TAA AAG AAT GGA G	28	57.8
	A0026IF2	ACA AGG GTA ATA ATT TTA AAA GAA TGG AG	29	58.2
	A0026IR1	GTT TTC ATA TTT TTG TTT CAT TTT TAT CC	29	55.4
	A0026IR2	AGT TTT CAT ATT TTT GTT TCA TTT T	25	49.9
	A0026OF1	TTC AAA ATA TGC CTC TTA AAC CCA ATA AGG	30	61.3
	A0026OF2	CAA AAT ATG CCT CTT AAA CCC AAT AAG G	28	60.7
	A0026OR1	GGA TTC GCT TGA GCA TGC TAG TAG	24	62.7
	A0026OR2	AAT ATG GAT TCG CTT GAG CAT GCT AGT AG	29	63.9
BG0027 (Q662Y5)	G0027IF1	CAA GGA TAG TAA TTT TAA AAG AAT GGA G	28	57.8
	G0027IF2	GGA TAG TAA TTT TAA AAG AAT GGA GA	26	55.3
	G0027IR1	GTT TTC ATA TTT TGT TTC ATT TTG TC	26	53.8
	G0027IR2	TTC ATA TTT TTG TTT CAT TTT TGT C	25	51.5
	G0027OF1	GAT TCT CAA TAT AAG AAA AGC TAA GCT AC	29	59.6
	G0027OF2	GGA TAG TAA TTT TAA AAG AAT GGA GA	30	59.9
	G0027OR1	CCT CTT GGA TTG AAC TGC CTA AAA TG	26	61.6
	G0027OR2	CTA GTT GGG AAA ATA TGG ATT CGC TTG AGC	30	65.4
BAPKO_0422	A0422IF1	TTG AAT ATA AAA AAG GAG AAT TTC A	25	51.5
	A0422IF2	ATA AAA AAG GAG AAT TTC AAT GAG	24	52.5
	A0422IR1	TTT AAA ATT TTT TAT CAT TTA TTC TCCA T	29	52.5
	A0422IR2	TTT ATC ATT TAT TCT CCA T	19	43.7
	A0422OF1	ATA CAG TCA AGT AGT ATA ATA ATA CCT GT	29	61.2
	A0422OF2	ACA GTC AAG TAG TAT AAT AAT ACC TGT AT	29	61.1
	A0422OR1	GAT TGC CTA TTC CAA TAC CCA TTG	24	61.3
	A0422OR2	GGA TTG CCT ATT CCA ATA CCC A	22	60.7
BG0407	G0407IF1	TAT AAA AAA GGA GAA TTT CAA TGA GA	26	53.8
	G0407IF2	GGA GAA TTT CAA TGA GAA TGC	21	54.0
	G0407IR1	TTT TAA AAT TTT TTA TCA TTT ATT CTC CAC	30	54.5
	G0407IR2	CCA CTA TAT ATA TAT TTT GAT AAA GCC TG	29	58.2
	G0407OF1	AAG ATT AAC ATC AAT ACC CAG TCA AGT AGT	30	61.3
	G0407OF2	ACC CAG TCA AGT AGT ATA ATA ATA CCT AT	29	59.6
	G0407OR1	GAT GCA AAA GAA AGG TGT GCT ACA CC	26	63.2
	G0407OR2	CTT CTT CCT TGC TGC ATC TGA CC	23	62.4

**Table 2.5. Primers for the amplification of target genes.** The primers were named according to the species of bacteria – A for *Borrelia afzelii* and G for *Borrelia garinii*, followed by the locus name – e.g. 0407, after which is the stage of nPCR – Inner or Outer, then the direction the primer will amplify – Forward or Reverse. In each instance, multiple primers were designed to screen for optimal combinations. This is indicated by the final number in the primer designation.

Restriction enzymes to be used in cloning were selected to be compatible in sequential digests. Before primer design the genomic sequence was checked to ensure the recognition sequence of chosen enzymes was not present at any point in the gene. The gene sequences to be targeted were run through a web server (NEBcutter v2.0) on the New England Biolabs website to determine which recognition sites were naturally present in the genetic sequence (Figure 2.1).



**Figure 2.1. Restriction sites present throughout the gene BAPKO\_0422.** Graphical output from NEBCutter v2.0. Each enzyme present in the figure above digested the expected genomic sequence at some point, making it unsuitable for use in the project.

From analysis of the genetic sequences to be targeted, *BamHI* and *HindIII* were chosen for use in restriction digestion as the restriction sites were not present at any point in the gene. Primers were designed for the addition of these restriction sites to PCR product (Table 2.6).

Target	Primer	Sequence (5'-3')	Length (bp)	T <sub>m</sub> (°C)
BAPKO_0026	A0026CF	GCA TGG ATC CGC ACG CAG TTA TTG ATA GA	29	50.71
	A0026CR	ATC GAA GCT TCT ACA ATT TAT TTA CAT TTA CTG	33	45.71
BG0027	G0027CF	GCA TGG ATC CGG CAG TCA ACA TTG ATA	27	45.60
	G0027CR	ATC GAA GCT TCT ACA ATT TAT TTA CAT TTA CT	32	42.19
BAPKO_0422	A0422CF	GCA TGG ATC CGC AAT CAA AAA GCA AAA CTA T	31	52.67
	A0422CR	ATC GAA GCT TTC ATT TAT TCT CCA TTA TAT ATA	33	44.85
BG0407	G0407CF	GCA TGG ATC CGA CCA ATC AAA AAG CAA ATC T	31	54.57
	G0407CR	ATC GAA GCT TCT ATA TAT ATA TTT TGA TAA AGC C	34	44.82

**Table 2.6. Primers for the addition of restriction sites.** Bases highlighted in red show the restriction sites. The four bases at the beginning of each primer were added to increase restriction enzyme efficiency. The calculated T<sub>m</sub> does not include the bases for digestion or those prior. T<sub>m</sub> calculations carried out using the Applied Biosystems T<sub>m</sub> calculator (<http://www6.appliedbiosystems.com/support/techtools/calc/>).

### 2.1.3 Plasmid vector

For cloning of genes, the plasmid pET-47b(+) was selected. pET-47b(+) was chosen for a variety of factors, including the presence of; a T7 promoter site, N-terminal histidine tag and a HRV-3C cleavage site (Figure 2.2 & Figure 2.3). pET-47b(+) also has a kanamycin resistance gene for selective growth.

The T7 RNA polymerase gene only allows transcription of DNA downstream with a low error rate; T7 RNA polymerase is highly selective for its promoter sequence and synthesises RNA many times more efficiently than *E. coli* RNA polymerase (Tabor, 2001). Induction by



## **2.2 Methods**

### **2.2.1 Nested PCR amplification and cloning of targets**

A master mix containing all reagents, with the exception of DNA and both primers was prepared before each set of PCRs. This master mix contained enough reagents for all the planned PCRs plus one extra to allow for unavoidable pipetting errors. Each PCR master mix contained: 1 unit Taq polymerase, 10 µl reaction buffer (1x), 2 mM MgCl<sub>2</sub> and 0.25 mM of each dNTP. This was added to tubes containing 2 µl of DNA (or initial reaction) and 2 µM of each primer. The DNA used for amplification was that supplied by Dr Gabi Margos, this was diluted 1:5 in nuclease free water prior to amplification.

The first stage of nPCR uses template gDNA and a set of primers encompassing a target approximately 150-300 base pairs larger than the fragment. This followed the protocol laid out for twenty cycles. The second stage of nPCR uses 2 µl of reaction from the initial amplification as a template and primer pairs much closer to (and sometimes overlapping) the target gene. The heating procedure for the first step was applied again, with thirty five cycles used. The annealing temperature was again adjusted based on the  $T_m$  of the primers used. Initially nested PCR was used to amplify the UvrA/PepX housekeeping genes from *B. garinii*.

Once the amplification of the target genes was complete, the genes were purified using a Qiaquick PCR Purification kit purchased from Qiagen (Manchester, UK). PCR using overhang primers designed for the specific restriction enzymes BamHI and HindIII was carried out on the targets in preparation for restriction digests.

### **2.2.2 Restriction digest and production of DNA constructs**

The restriction enzymes BamHI and HindIII were used for digesting pET-47b(+) and prepared genes. Following an enzyme compatibility chart on the New England Biolabs website (New England Biolabs) the restriction digests were carried out sequentially. The reaction mix for the initial digest consisted of 3 µl DNA, 1 µl neat enzyme, 2 µl of appropriate NEBuffer and nuclease free water up to a total volume of 10 µl. The second digest reaction mix consisted of: the 10 µl first reaction, 1 µl of enzyme, 4 µl NEBuffer, 1x (40 ng) BSA (required for BamHI digests) and nuclease free water up to a total volume of 20 µl. Each restriction digest was incubated at 37 °C for one hour with gentle shaking (120 rpm).

Once the digest was complete the DNA was purified using a QIAquick PCR Purification kit (Qiagen, Crawley, UK) and eluted in 30 µl nuclease free water. This produced pure insert/plasmid DNA ready for cloning. In addition to removing any enzymes and buffer components present in the samples, the purification step also removes small unwanted fragments of DNA, for example the extra bases next to each restriction site and the small fragment produced by the digestion of pET-47b(+).

DNA concentrations were determined using the absorbance at 260 nm and the equation below (equation 1). The number 50 is used as a DNA concentration of 50 µg/ml produces an absorbance of 1 at 260nm.

$$\text{Concentration (}\mu\text{g/ml)} = (A_{260} - A_{320}) \times 50 \times \text{dilution factor}$$

**Equation 1. Calculating the concentration of DNA.**

The  $A_{280}$  was then used in conjunction with the  $A_{260}$  to calculate the purity of the DNA sample as shown below (equation 2). For the sample to be of an acceptable purity, the ratio had to be above 1.5.

$$\text{Sample purity} = \frac{A_{260}}{A_{280}}$$

**Equation 2. Calculating DNA purity.**

To have a large enough sample volume for use in a low capacity cuvette (70 µl capacity) the sample was diluted by a factor of 3. Both the concentration and the purity ratios were calculated.

The volumes required for ligation were determined by: calculating the volume of plasmid DNA required to give 1 ng of DNA, then calculating the difference in size of the insert and plasmid as a ratio, before using that to deduce the volume of insert DNA to use.

The ligation reaction protocol went through several iterations before one was found that produced gene constructs. The concentration of ligase was increased 10 fold (to the neat solution), and the ATP concentration was increased to 2 mM (from 1 mM in the initial buffer). These incorporated changes were not successful either in conjunction or separately. Different ligation temperatures and incubation times were trialled. 23 °C for one hour with shaking at 120 rpm produced DNA constructs. The final ligation mix used pET-47b(+):gene

ratios of 1:1, 1:3, 1:6 and 1:10, mixed with 1 U T4 ligase, 1 mM ATP, 1 µl ligation buffer, with the volume being made up to 10 µl total with nuclease free water. This was incubated at 23 °C for one hour at 120 rpm.

### **2.2.3 Agarose gels.**

Products of the PCR amplifications and restriction digests were run on an agarose gel to determine a) the success of the amplification and b) success of the digestions. 1.5 % (w/v) agarose gels were prepared by melting 1.5 g agarose in 200 ml 0.5x TBE in a 250 ml conical flask, to which 20 µl SYBR safe DNA stain was added once the flask was cool enough to touch. This was set and once loaded, run for 4 hours at 125 kV and 200 mA in 0.5 x TBE buffer. As PCR amplifications were conducted in Green GoTaq Reaction Buffer (Promega), the crude samples were ready to load. Purified samples (after restriction digest) had 6x Blue/Orange loading dye added in the required ratios, as did purified digested plasmid. The DNA ladder was chosen depending on the size of products expected; if plasmids were involved, both ladders (max 800 and 10,000 bp) were run, if the gel was for amplified genes only, only the 800 bp ladder was run. The 10,000 bp ladder was provided ready to use, whereas the 800 bp ladder (2 µl) required Blue/Orange loading dye (2 µl) and nuclease free water (8 µl) to prepare. If greater resolution of bands was required, the gel was run for 6 hours at 90 kV, 200 mA.

### **2.2.4 Competent cell preparation**

The *E. coli* strain chosen for plasmid maintenance was NEB Turbo Competent *E. coli* (High efficiency) from New England Biolabs (NEB, Hitchin, UK). 50 µl of cells were spread onto an agar plate with no antibiotic and incubated at 37 °C overnight. Of the colonies that grew, one was selected, transferred to 5 ml sterile LB and incubated at 37 °C overnight. This culture was transferred to 50 ml sterile LB, incubated at 37°C with gentle (120 rpm) shaking and the OD<sub>600</sub> monitored until it reached 0.4, upon which the culture was split and incubated on ice for 5 minutes before being centrifuged at 10,000 g for 30 minutes. Each pellet was resuspended in 2.5 ml chilled TSS buffer and 20 µl aliquots frozen -70 °C.

### **2.2.5 Transformations**

20 µl aliquots of competent cells were thawed on ice, and either purified ligation reaction mix or 10 ng construct DNA added. Mixing was induced by gentle flicking of the tube and the

cells incubated on ice for 15 minutes before being transferred to a 42 °C water bath for a 45 second heat shock, after which they were returned to ice for 5 minutes. 80 µl SOC was added and the cells were incubated at 37 °C for one hour. The entire volume of cells was spread on a pre-warmed agar selection plate (containing appropriate antibiotic) and incubated at 37°C overnight

Control transformations were carried out at all times:

- Positive control 1 – unaltered plasmid DNA
- Positive control 2 – plasmid DNA digested with a single restriction endonuclease and re-ligated
- Negative control 1 – No DNA
- Negative control 2 – plasmid DNA digested with a single restriction endonuclease (no DNA ligase added)
- Negative control 3 – plasmid DNA digested with both restriction endonucleases (no DNA ligase added)

### **2.2.6 Extraction of plasmid constructs**

From cells cultured on LB-agar-antibiotic plates, one of the colony was selected and cultured overnight in 5 ml sterile LB with appropriate antibiotic(s) at 37 °C with gentle shaking. The following morning construct DNA was extracted using a QIAprep Spin Miniprep Kit (Qiagen, Crawley, UK) according to the manufacturer's protocol and eluted in 50 µl nuclease free water before being stored at -20°C.

Four different strains of competent expression *E. coli* were screened for best production of protein. The three strains selected were:

- T7 Competent *E. Coli* (High efficiency) (NEB, Hitchin, UK)
- *E. coli* BL21 (University stock)
- *E. coli* BL21 (DE3) (NEB, Hitchin, UK)
- *E. coli* Rosetta (DE3)

Stocks of all four strains were prepared using the same procedure as used for preparing the maintenance strain.



## **2.2.7 Expression of protein in inclusion bodies**

### **2.2.7.1 Expression of protein using IPTG induction**

One colony from the selection-plates was selected and sequentially cultured at 37 °C in 5 ml, 50 ml and (and in the case of larger-scale expression, 500 ml) LB with antibiotic, with the OD<sub>600</sub> being monitored in both the 50 and 500 ml cultures until it reached 0.6. Once the critical OD<sub>600</sub> was reached a 1 ml sample was taken for pre-expression SDS-PAGE analysis. To induce expression IPTG was added to a working concentration of either 1 or 0.1 mM and the cultures incubated at 37 °C with vigorous shaking for three to four hours, upon completion a further 1 ml sample was taken for post expression SDS-PAGE analysis. Other expression conditions included expression at 18 and 20 °C overnight.

### **2.2.7.2 Expression by autoinduction**

Autoinduction media was prepared following the recipe detailed in the buffers and media (Table 2.2). The base media containing tryptone and yeast extract was autoclaved before sterile NPS, 5052 and antibiotic (0.45 µM filtered) were added to the media. These components had to be added after autoclaving to prevent the sugars caramelising and compounds degrading. A small volume of construct-containing expression cells was introduced to the media (~1 ml) and the culture incubated at 37 °C for between 18 and 24 hours. Initially the cultures were shaken at 120 rpm, after experimentation this was increased to 200 rpm.

## **2.2.8 SDS-PAGE and Western Blot analysis**

### **2.2.8.1 Preparation of samples**

Whole cell samples to be run on an SDS-PAGE were harvested by centrifugation for 5 minutes at 14,000 rpm before being resuspended in 100 µl of either salt or urea buffer. The samples were then sonicated for a total of 30 seconds (4 second pulse, 5 second rest, 75 % amplitude). Once the samples were ready, a small volume (5 – 10 µl) was taken, SLS added to a working concentration of 1x, along with 3 µl 1 M DTT the volume made up to 25 µl with ultrapure water and the sample boiled at 95 °C for 5 minutes. Protein samples containing guanidine-HCl to be analysed by SDS-PAGE and Western Blot had to be extracted from the denaturant due to guanidine precipitating upon mixing with SDS. To each sample 9 volumes of 100 % (v/v) ethanol was added and the solution incubated at -20 °C for one hour. Protein was harvested by centrifugation at 14,000 rpm for fifteen minutes in a bench-top centrifuge

and the pellet resuspended in 100 µl 70 % (v/v) ethanol before being centrifuged for a further five minutes at 14,000 rpm. The final pellet was resuspended in 100 µl of either salt buffer or 8 M urea.

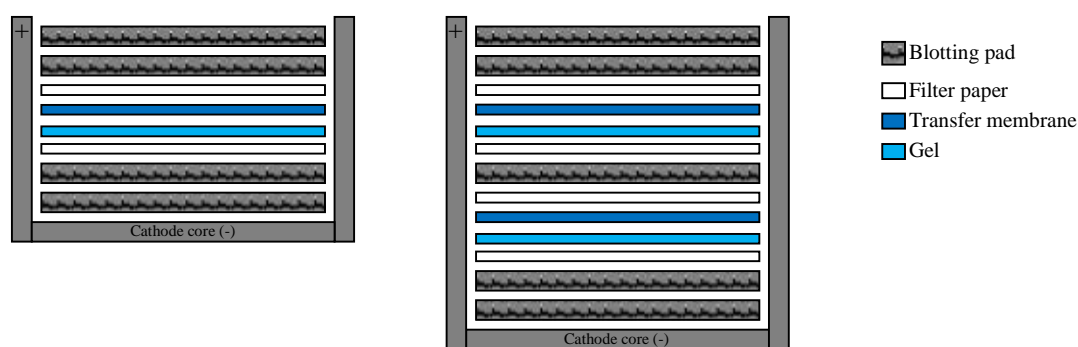
#### **2.2.8.2 SDS-PAGE & Western Blot**

Both pre-cast and in-house cast gels were used throughout this research. 4-12 % (w/v) Bis/Tris pre-cast gels were purchased from Invitrogen (Paisley, UK), 10 % (w/v) Bis/Tris gels were cast using a water cooled PerfectBlue Dual Gel Twin S gel tank (Peqlab, Sarisbury Green, UK). A 10 ml gel was prepared using the following recipe: 7.35 ml ultrapure water, 3.75 ml 1 M Tris-HCl (pH 8.8), 75 µl 20 % (w/v) SDS, 3.75 ml Bis/Tris acrylamide (40 % w/v), 75 µl 10 % (w/v) APS and 10 µl TEMED. The stacking buffer was prepared as follows: 3.25 ml ultrapure water, 1.25 ml 1 M Tris-HCl (pH 8.8), 25 µl 20 % (w/v) SDS, 0.5 ml Bis/Tris acrylamide (40 % w/v), 25 µl 10 % (w/v) APS and 5 µl TEMED. 20 µl of protein sample was loaded into the wells. 7.5 µl Novex Sharp Pre-stained protein standards (Invitrogen, Paisley, UK) was run for mw comparison with each gel. Pre-cast gels were run at 200 kV, 400 mA for 35 – 40 minutes. In-house cast gels were run at 175 kV, 400 mA until the loading dye reached the bottom of the gel. All gels were stained for one hour using Coomassie Blue stain before destaining over the course of several hours. It is important to note that due to the presence of hydrophobic regions in membrane proteins there is the potential for anomalous migration on an SDS-PAGE; either due to the protein binding excessive amounts of SDS or the improper unfolding of hydrophobic regions (Hunte *et al.*, 2003).

Western blotting was carried out using an XCell II Blot module (Invitrogen, Paisley, UK). The primary antibody was selected to be specific for the his-tag; a mouse monoclonal anti-his (IgG<sub>1</sub>-kappa isotype). The secondary antibody used was a goat anti-mouse polyclonal Alexa Fluor 680 antibody. Both antibodies were purchased from Invitrogen (Paisley, UK). Prior to use the primary antibody was diluted 1:1000 (1 µg/ml) in PBS-TWEEN and the secondary antibody diluted 1:5000 (1 µg/ml).

In preparation for Western blotting, four blotting pads and two sheets of filter paper cut to the correct size were soaked in transfer buffer for several minutes. The PVDF transfer membrane was pre-wet in methanol for 30 seconds before being rinsed with dH<sub>2</sub>O and soaked in transfer

buffer. Once electrophoresis was complete, the gel had the wells and foot removed before the membrane was placed on one side and the pair sandwiched between two pieces of filter paper. This ‘gel-membrane sandwich’ was placed on top of two blotting pads in a blotting module, taking care to ensure that the gel was underneath the membrane and closer to the cathode. Another two blotting pads were placed on top as shown below (Figure 2.4). This was transferred to the running tank and the middle chamber was filled with transfer buffer, whilst the outer chamber was filled with deionised water. The running tank was surrounded by ice to keep the apparatus cool. Blotting was run for 90 minutes at 30 V, 400 mA.



**Figure 2.4. Western blot module set up. A) Single membrane transfer, B) transfer of two membranes simultaneously**

Once the transfer was complete, the membrane was removed and washed for one hour in blocking solution (2 % w/v milk powder in PBS) at 4 °C. The membrane was then probed with primary antibody for one hour at 4 °C before being probed with the secondary antibody for a further hour at 4 °C. Between each step the membrane was washed three times with PBS-TWEEN for 5 minutes each.

Once the final wash had been completed, the membrane was scanned using a LI-COR Odyssey infrared imaging system (LI-COR Biosciences) at a wavelength of 680 nm. The intensity, brightness and contrast were adjusted to give the best view of the bands present.

### **2.2.9 Purification of inclusion bodies from autoinduction cultures.**

Cell pellets harvested from expression were resuspended in a volume of lysis buffer 1/20<sup>th</sup> the original volume and incubated on ice for one hour before sonication for a total of 1 minute (11 second pulse, 49 second rest, 95 % amplitude), after which the insoluble fraction was separated by centrifugation at 12,000 g for 20 minutes at 4 °C. The supernatant containing soluble proteins and cellular debris was poured off and stored at – 20 °C. The insoluble pellet

was resuspended in salt buffer and centrifuged again and the supernatant poured off, washing the pellet. This was repeated once. After the pellet had been washed it was resuspended in FPLC Buffer A (containing 6 M GuHCl) and centrifuged for 1 hour at 25,000 g, 4 °C. A protease inhibitor cocktail (containing no EDTA) (Sigma) was added to the supernatant (now containing insoluble proteins) which was stored at -20 °C. Initially all denaturing buffer systems were urea based (8 M), however, once the purification was switched to using 6 M GuHCl, the resolubilisation buffer was too.

## **2.2.10 Immobilised Metal Affinity Chromatography**

Immobilised Metal Affinity Chromatography (IMAC) is usually carried out using buffers with pH between 6 and 8.5 depending on the protein. Optimal separation of protein generally occurs in the presence of 0.1 – 1 M NaCl, this is known to reduce the potential for membrane protein aggregation in addition to eliminating ion exchange effects and improve solubility. Detergents are required to solubilise the protein, neutral detergents don't interfere with binding and elution and a wide variety of buffer conditions can be used. Elution is carried out by competitive elution through a reduction in pH or the addition of imidazole (Hunte *et al.*, 2003). Aggregation is a factor when working with membrane proteins as it occurs even in the presence of detergents, reducing the efficiency of all separation techniques (Hunte *et al.*, 2003). This is accounted for by optimising the pH for solubilisation and stability, and optimisation of detergent concentrations.

IMAC was carried out on an Äkta Prime FPLC (GE Healthcare, Amersham, UK) using a 5 ml HisTrap HP Ni-NTA column (GE Healthcare, Amersham, UK) and a variety of buffers, dependant on whether on-column refolding was being carried out. A three buffer method was developed for the purification of insoluble BAPKO\_0422 and a seven buffer system was developed for the purification and on-column refolding. All buffers were made fresh before each purification and were vacuum filtered using a 0.45 µm filter.

### **2.2.10.1 Purification of insoluble BAPKO\_0422**

The purification system and column were equilibrated in the buffers to be used (a minimum of 6 column volumes for the column), the pipelines were washed until the absorbance and conductivity baselines stabilised. The sample was loaded onto the column and FPLC Buffer A pumped through to wash off any unbound proteins until the absorbance returned to the

baseline. FPLC Wash Buffer containing 20 mM imidazole was pumped through the column to elute any proteins that had weakly bound to the Ni-NTA due to exposed histidine residues, until the baseline again returned to zero. Bound proteins were eluted using gradient elution over either 50 or 100 ml with the concentration of FPLC Buffer B (containing 0.3 M imidazole) increasing from 0 to 100 % over this time.

#### **2.2.10.2      Benchtop refolding of BAPKO\_0422**

BAPKO\_0422 was adjusted to a concentration of 1 mg/ml using 0.1 % (w/v) LDAO, 0.3 M NaCl and 50 mM Tris (pH 7) and added over a period of ten minutes to a volume of SEC buffer 50 x larger than the protein volume (McConnell and Pachón, 2011). The protein solution was left to refold at room temperature whilst being continuously stirred using a stirrer hotplate for one hour. The dilute refolded protein was concentrated first by IMAC purification using non-denaturing buffers before being concentrated in a 10,000 mwco concentrator.

#### **2.2.10.3      On-column refolding of BAPKO\_0422**

For on-column refolding of BAPKO\_0422, the preparation and protein loading procedure is identical to that detailed in section 2.2.10.1 up to and including the low-concentration imidazole wash. Once this was complete, four buffers were used to lower the concentration of GuHCl and introduce LDAO to the protein, this reduction in denaturant in the presence of 0.1% (w/v) LDAO would allow the protein to refold into detergent micelles; the CMC of LDAO is 0.023 % (w/v) (Herrmann, 1966). RF1 (4 M GuHCl) was pumped through at a flow rate of 2 ml/min for three column volumes (15 ml), after which the instrument paused and RF2 (2 M GuHCl) pumped through for the same volume at the same flow rate. This was repeated with buffers RF3 (1 M GuHCl) and RF4. The final refold buffer contained no GuHCl and is the same buffer used for size exclusion chromatography. The protein was then eluted by gradient elution from RF4 to elution buffer over the course of 50 ml and purity assessed by SDS-PAGE.

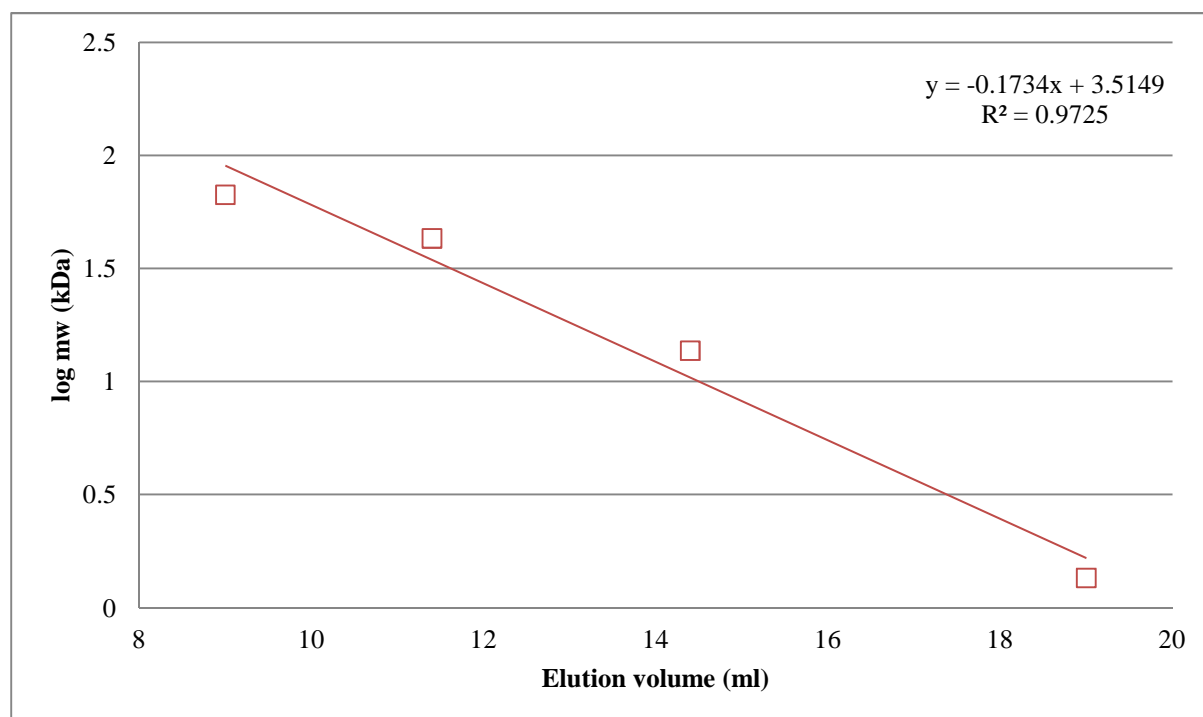
#### **2.2.11          Size Exclusion Chromatography**

Size exclusion chromatography (SEC) was used to further purify the protein and provide an initial estimation for the size of molecule produced. When purifying membrane proteins using SEC it is important to keep the concentration of detergent above the CMC but below 5

% (w/v), and the protein concentration below 10 mg/ml. Detergent bound to the protein can increase molecular weight (eg Triton X-100 micelles add 90 kDa to the molecular weight) (Hunte *et al.*, 2003). SEC was carried out using a Superdex™ 75 10/300 GL column (GE Healthcare), which was stored in 20 % (v/v) ethanol when not in use. The column was washed through with water to remove all traces of ethanol, before being equilibrated with SEC buffer. The SEC buffer for BAPKO\_0422 consisted of 0.5 M NaCl, 50 mM Tris, 0.1 % (w/v) LDAO at pH 8.0. A calibration curve (Figure 2.5) was produced using four protein standards:

- 0.5 ml BSA (67 kDa) 8 mg/ml
- 0.5 ml Ovalbumin (43 kDa) 2.5 mg/ml
- 0.5 ml Ribonuclease (13.7 kDa) A 5 mg/ml
- 0.5 ml Vitamin B12 (1.35 kDa) 12 mg/ml

Standards were selected based on the manufacturers specifications. The elution times agreed approximately with those provided from the manufacturer.



**Figure 2.5. SEC calibration curve.** Produced using four protein standards as detailed above. BSA eluted first, followed by ovalbumin, ribonuclease A and finally vitamin B12.

Between 0.3 and 0.5 ml concentrated BAPKO\_0422 was manually loaded onto the column using a 1 ml syringe. The absorbance of the flow through was monitored at 280 nm throughout. Small membrane proteins bind more detergent per gram than larger proteins, masking differences in size, mass and moiety. Therefore techniques based around differentiation of molecular mass and sizes (such as SEC) have lower resolution (Hunte *et al.*, 2003).

### 2.2.12 SEC-MALLS

Size-exclusion chromatography linked with multi-angle laser light scattering (SEC-MALLS) was used to estimate the molecular weight of protein in the protein-detergent micelle. SEC-MALLS was carried out using Astra software version 5.3.4.20 monitoring three conditions:

- Differential refractive index (Optliab rEX detector)
- Light scattering (DAWN EOS instrument)
  - Cell type - K5
  - Laser wavelength - 690 nm
  - Calibration constant -  $6.0520 \times 10^{-6} \text{ 1/(V cm)}$
- Ultraviolet absorbance (Providence UV/Vis SPD-20A)
  - Wavelength 280 nm

Three Agilent PL Aquagel-OH (40, 50 and 60) SEC columns (1.5  $\mu\text{m}$  id) were linked in sequence before being linked in series to the LS and UV detectors. The system was equilibrated using SEC buffer at 0.45 ml/min until the light scattering baseline variance was acceptable.

150  $\mu\text{l}$  8 mg/ml BSA was run as a positive control. Analysis of the data assuming a  $\text{dn/dc}$  value of 0.187 ml/g gave a molecular weight of 60 kDa  $\pm$  15 % without manual background adjustment. As the molecular weight of monomeric BSA is 66 kDa, this was deemed acceptable. A 300  $\mu\text{l}$  BAPKO\_0422 sample of approximately 5 mg/ml was used for analysis. After the run had completed the baselines were manually adjusted, the peak manually selected and the molecular weight protein of the protein calculated by the software using a standardised  $\text{dn/dc}$  value of 0.187 ml/g (Slotboom *et al.*, 2008).

### 2.2.13 Circular Dichroism

Two criteria should be met to obtain a good CD spectra: the total absorbance of the sample should remain below 1.0 and the High Tension (H[T]) voltage should not exceed acceptable parameters (instrument variable) as this decreases the reliability of the results (Kelly *et al.*, 2005). The main processes of meeting these criteria are altering the sample concentration or the cell path length. The signal/noise ratio can be lowered by conducting repeat measurements. For the JASCO J-810 instrument that was used in this research, the maximum acceptable H[T] value is 600 V, above which the CD spectra generated becomes unreliable.

The percentage secondary structure composition (helix, strand, turn or unordered) of proteins can be determined from the CD due to absorption of polarised light by peptide bonds at wavelengths below 240 nm, aromatic side chains between 260 and 320 nm and disulphide bonds at around 260 nm (Kelly *et al.*, 2005). Different secondary structures exhibit characteristic peaks at certain wavelengths:  $\alpha$ -helical proteins have negative peaks at 208 and 222 nm and a positive peak at 193 nm and proteins containing  $\beta$ -sheets have a positive peak at 195 nm and a negative peak at 218 nm (Greenfield, 2006). Calculation of the secondary structure proportions is done by a variety of algorithms such as SELCON, CONTIN and CDSSTR. The online server DichroWeb (Lobley *et al.*, 2002, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008) uses these along with a variety of reference CD spectra to calculate the secondary structure composition (Janes, 2008).

Circular dichroism was carried out at the University of York on a Jasco J-810 Spectropolarimeter with the assistance of Dr Andrew Leech. An air background spectra as initially taken, followed by the buffer background spectra in triplicate. Measurements were taken over a wavelength range of 190 – 260 nm using a cuvette with a path length of 0.2 mm. Samples were prepared for circular dichroism by dilution in the SEC buffer to approximate concentrations of 0.166 and 0.33 mg/ml. The CD signal was measured at various temperatures from 260 to 190 nm. All experimental data was acquired in triplicate.

Data obtained were analysed using the DichroWeb online CD server (Lobley *et al.*, 2002, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008) using the following parameters:

- File format: JASCO 1.30
- Input units: millidegrees / theta (machine units)



- Initial wavelength: 260 nm
- Final wavelength: 190 nm
- Wavelength step: 0.5 nm
- Lowest data point to use: 190 (DichroWeb would not accept higher to discount unreliable data as the reference datasets were optimised to 190 nm)
- Analysis program: CDSSTR or SELCON3
- Reference set: 4 or 7 (both optimised for 190 – 240 nm)
- Output units: mean residue ellipticity
- Mean Residue weight (Da): 112
  - Calculated by: mw (inc his-tag) / no. of residues – 1
- Protein concentration: 0.33 mg/ml or 0.166 mg/ml
- Path length: 0.2 cm.

## **2.2.14 Small Angle X-Ray Scattering**

### **2.2.14.1 Collection of data**

Background data sets were collected for a period of 24,000 seconds (approximately 3 hours, 40 minutes) and 65,000 seconds (18 hours) on a Bruker NANOSTAR. Scattering data for the following buffers was acquired:

- SAXS Salt Buffer (SSB) - 0.3 M NaCl, 50 mM Tris, pH 8.0
- SAXS LDAO buffers containing 0.01 % (w/v), 0.05 % (w/v) and 0.1 % (w/v) LDAO in addition to the salts in SSB
  - Low composition SAXS LDAO buffer - 0.01 % (w/v) LDAO, 50 mM NaCl, 20 mM Tris-HCl, pH 8.0
- SAXS CHAPS buffers containing 0.01 % (w/v), 0.05 % (w/v) and 0.1 % (w/v) CHAPS in addition to the salts in SSB

The CMC of LDAO, like all detergents varies with salt concentration, values found in the literature quote approximately 0.046 % (w/v) (SIBYLS Beamline) to 0.023 % (w/v) in water and approximately an eighth of this in 0.1 M NaCl (Herrmann, 1966). Hence the LDAO concentration was expected to be above the CMC.

Scattering data was acquired for lysozyme (10 mg/ml) and BAPKO\_0422 (6 and 10 mg/ml ) samples in SAXS LDAO Buffer containing 0.1 % (w/v) LDAO using the same method of collection as for background data. Acquisition times were 65,000 seconds. BAPKO\_0422 data was collected at two different concentrations to provide information about the state of the sample. Lysozyme was used as a positive control, and in conjunction with the collection of background detergent data sets this ensured that protein scattering could be detected above detergent scattering, whilst providing a valuable example of data analysis. Background subtraction was conducted using water and buffer background scattering using the software provided.

### 2.2.14.2 Analysis of data

PRIMUS is a data manipulation program with an automatic function for producing Guinier plots and evaluating parameters from Guinier and Porod plots using SAXS data sets (Konarev *et al.*, 2003). The raw data file is input into the program and automatically plotted. The Rg can also be calculated automatically.

There are several parameters to be set when running GNOM. Most were left as the default setting, however some were adjusted and are listed below.

- Angular scale – 2. units of  $4 \cdot \sin(\ )$  (units that the data uses -  $\text{nm}^{-1}$ )
- Type of system – 0. calculation of distance distribution function for a monodisperse system. The function  $P(r) = (r) \cdot r^2$  is evaluated, where  $(r)$  is the characteristic function of the particle.
- Zero condition at  $r=R_{\min}$  – yes.
- Zero condition at  $r=R_{\max}$  – yes.
- $R_{\max}$  for evaluating  $P(r)$  – 50. expected maximum size of particle ( $\text{\AA}$ )

DAMMIF is a high-speed version of DAMMIN (25-40 times faster) that performs shape determination from small angle X-ray data *ab initio* by simulated annealing using a single phase dummy atom model (Franke and Svergun, 2009). A particle is represented as a collection of a large number of beads, each of which belongs to the molecule or the solvent. This program uses the output file from GNOM

- Angular unit – n (data is in nm)
- Mode – slow (fast is used to get a preliminary idea of the molecular envelope)

- Anisometry – prolate

The program then builds a model of the molecular envelope using small spheres and measuring the fit of the data after each step. The lower the fit, the better match the model.

DAMMIF was run ten times to produce a set of molecular envelopes for averaging. These molecular envelopes were read into DAMSEL which aligned and compared the models using Normalised Spatial Discrepancy (NSD) values to find the most probable real-world solution; if the NSD for two models was less than 1, they were considered similar. The mean NSD and variance is automatically calculated and any solutions that were outside of acceptable limits (2 x variation either side of the mean) were discarded (Volkov and Svergun, 2003).

Solutions accepted by DAMSEL were then read into DAMSUP. This superimposes each of the molecular envelope structures onto each other using SUPCOMB, producing models of the superimposition (Volkov and Svergun, 2003).

DAMAVAR, takes the models aligned by DAMSUP and averages them to create a probability map (aligns the solutions to give the best overlap). This remaps the .pdb files onto a grid of densely packed beads to compute a frequency map, which is then processed to provide a filtered model using DAMFILT. It should be noted that the DAMAVAR file will not be a good fit to the data (Volkov and Svergun, 2003).

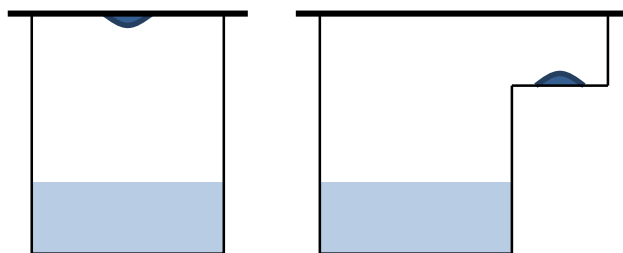
The DAMAVAR averaged file was then processed by DAMFILT. This filters the DAMAVAR model at a given cutoff and removes low occupancy and loosely connected atoms, producing a compact and most probable core model (Volkov and Svergun, 2003). This is again not expected to fit the data, but is instead used as a starting point for DAMMIN to produce a molecular envelope using this and the experimental data. All parameters are left as default when producing, refining and averaging molecular envelopes.

The fit of the expected scattering patterns of the molecular envelope solutions generated by DAMAVAR, DAMFILT and DAMMIN were compared to the experimental data by CRY SOL and fits plotted (Svergun *et al.*, 1995).

## 2.2.15 X-Ray crystallography

### 2.2.15.1 *In surfo* crystallisation

There are various methods of crystallisation, in one of which the proteins (in this case detergent-solubilised membrane proteins) are utilised directly in crystal trials. *In surfo* methods include vapour diffusion and batch-phase (or micro-batch) crystallography (Cherezov *et al.*, 2006). Vapour diffusion crystallography takes place in 24, 48 or 96 well trays. In the bottom of each well is a reservoir of precipitation solution and buffer, or ‘mother liquor’. One drop of mother liqueur (from 0.1 – 0.5  $\mu$ l in 96 well trays to 1 - 2  $\mu$ l in 24 well trays) is added to similar volumes of protein (concentration varies depending upon solubility) in a sealed environment set up as either a hanging or sitting drop (Figure 2.6). The tray incubated at approximately 18 °C for a number of weeks. The concentration of precipitants in both the drop and the well gradually equilibrates over time, while water in the drop evaporates, steadily increasing the protein concentration and causing precipitation (Rhodes, 2000).



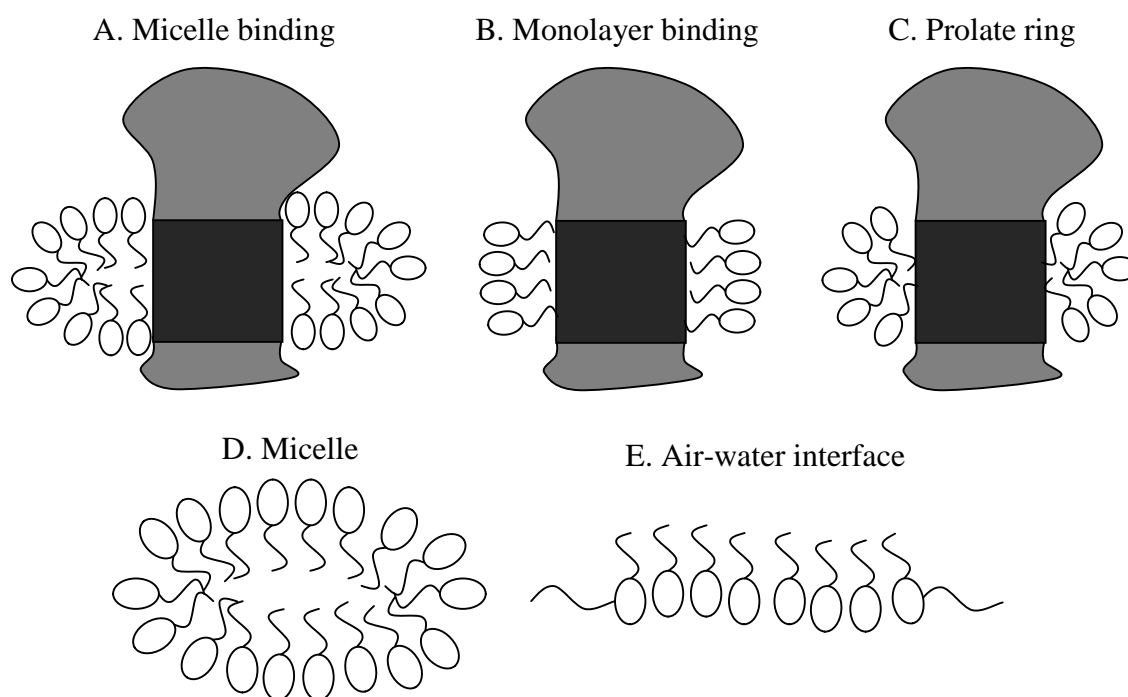
**Figure 2.6. Vapour diffusion setups.** Left - the hanging drop method, right – the sitting drop set-up. One major advantage of this method is the ease at which the drop can be accessed for harvesting crystals. The sitting drop techniques lends itself to 96 well automated setup in very small volumes and the entire tray is generally sealed with sealing film. The hanging drop method lends itself to 24 well screens and manual setups, with cover slips sealing each individual well.

In this research the hanging drop method was used in 24 well trays. In the case of CSS screens I and II, the reservoir was filled with 0.9 ml precipitant and 0.1 ml pH buffer (either 1 M Tris at pH 8 or 1 M sodium acetate at pH 4.5). For the Memb-PASS screen 0.4 ml precipitant was used. The protein and precipitant drops were mixed in various ratios on the surface of a cover slip, which was placed drop down on top of the well with a layer of vacuum grease in between as a seal. Trays observed under a microscope and notes taken with regards to the drops before being stored in a polystyrene box in a 20 °C incubator.

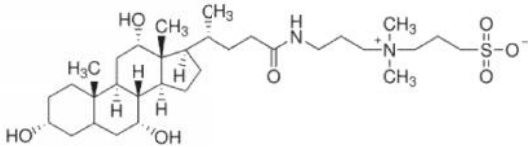
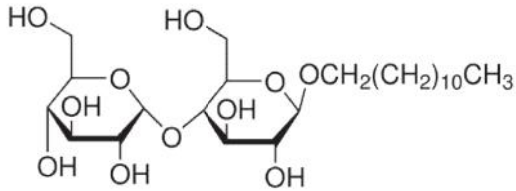
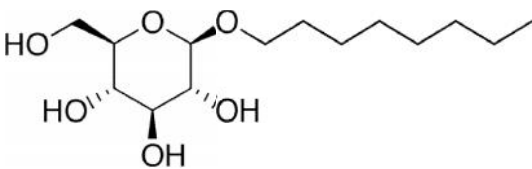
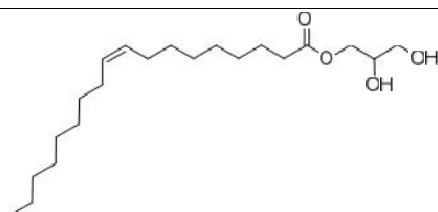
### 2.2.15.2 Detergents

Membrane proteins are challenging proteins to work with, both in crystallography and structural studies due to presence of the hydrophobic TM region. These regions can be stabilised through the use of detergents in various conformations (Figure 2.7) and/or lipids, thereby preventing aggregation.

The choice of which detergent to use is an important one and factors such as hydrophobic chain length, critical micelle concentration (CMC – the concentration at which the detergent spontaneously forms micelles) and cost must all be taken into consideration. A selection of detergents were chosen for use in this research based on a number of considerations. Table 2.7 shows detergents and lipid chosen for stabilising and crystallising the protein.



**Figure 2.7. Detergent binding by membrane proteins.** The dark grey section is the hydrophobic membrane-spanning region. **A)** shows a micellar arrangement of detergent molecules around the TM domain that has dimensions approximately the same as the detergent micelle. **B)** shows the TM domain is covered by detergent molecules in a layer that leaves the hydrophobic region of detergent exposed. **C)** represents an arrangement in which the hydrophobic regions of both the protein and detergent are shielded, similar to a micelle, but smaller. **D)** shows a micelle of pure detergent and **E)** represents detergent at an air-water interface. Diagram adapted from (LeMaire *et al.*, 2000).

Detergent	Structure	Mw (Da)	CMC (mM)	CMC (%)	Micelle size (kDa)
C <sub>8</sub> E <sub>4</sub>	$\text{CH}_3(\text{CH}_2)_6\text{CH}_2\left[\text{OCH}_2\text{CH}_2\right]_4\text{OH}$	306.44	8	0.25	25
CHAPS		614.9	8	0.49	6
DDM		510.62	0.12 - 0.17	0.0087	72
LDAO	$\text{CH}_3(\text{CH}_2)_{10}\text{CH}_2-\text{N}(\text{CH}_3)_2-\text{O}$	229.4	1-2	0.023 - 0.14	17 - 21.5
OG		292.4	18 - 30	0.53	7 - 29
Lipid - MO		356.54	-	-	-

**Table 2.7. Detergents and lipid commonly used in membrane protein purification and crystallisation.** Data for this table was obtained from various sources (Herrmann, 1966, Hjelmeland *et al.*, 1983, Chattopadhyay and London, 1984, VanAken *et al.*, 1986, Lorber *et al.*, 1990, Cortes and Perozo, 1997, LeMaire *et al.*, 2000, Lipfert *et al.*, 2007, Strop and Brunger, 2005)

LDAO was chosen as the primary detergent as it had previously been used to crystallise OmpA-like membrane proteins (Pautsch *et al.*, 1999, Pautsch and Schulz, 2000, Hong *et al.*, 2006) and due to the relatively small micelle size, the relatively low cost and low CMC. DDM, OG and C<sub>8</sub>E<sub>4</sub> were chosen as membrane proteins have been crystallised using these detergents previously (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Pautsch and Schulz, 2000, Albrecht *et al.*, 2006, Hong *et al.*, 2006). CHAPS was chosen as it is non-denaturing. MO was purchased for use in lipid cubic phase crystallography.

### **2.2.15.3     *In meso* crystallography**

Lipid cubic phase (LCP) crystallography (also known as *in meso* crystallography) utilises a stable lipid cubic phase as a basis for formation of protein crystals (Caffrey and Cherezov, 2012). The lipid cubic phase consists of monoolein hydrated with 40 % (w/w) water. This is formed by the combination of protein detergent solution at a concentration of 10 mg/ml with molten monoolein (MO) in a ratio of 2:3 (Caffrey, 2000, Caffrey, 2003, Cherezov *et al.*, 2006, Caffrey and Cherezov, 2012). Molten MO was prepared by dissolving 6 mg in an excess of chloroform (> 200 µl), before the majority of chloroform was evaporated by a nitrogen stream. The remainder was evaporated using a rotary evaporator over a period of four hours. Precipitant is added in excess (>10 parts with respect to the previous ratio) and incubated at 20 °C (Ai and Caffrey, 2000, Caffrey, 2000). The mixing of MO and protein solution forms the cubic phase spontaneously (Caffrey, 2000).

The initial LCP method used a piece of double-sided tape with holes punched in to form wells. A small volume of LCP with protein was deposited into each well. A 1 µl precipitant solution drop was applied to each LCP-protein deposit and the well sealed with a coverslip. This was carried out for all conditions in the Memb-PASS screen. The slides were incubated at 20 °C and checked after two weeks.

The second version of LCP crystallography used ‘wells’ made from a layer of grease on a microscope slide and covered with a cover slip.

## Chapter 3. Identification and bioinformatic studies of potential OmpA-like proteins in *Borrelia*.

Attempts were made to identify OmpA-like membrane spanning domains in *Borrelia* using BLAST searches of the sequences of OmpA (N-terminal domain), OmpW and OmpX proteins (1BXW, 2F1T and 1Q8J respectively) against the proteomes of *B. afzelii*, *B. burgdorferi* s.s., and *B. garinii* (BLAST search hosted on ExPASy (Altschul *et al.*, 1997)). An example of BLAST search results is shown (Table 3.1) using the N-terminal domain of OmpA from *E. coli* (1BXW) searched against the *B. garinii* proteome.

Accession no.	Description	Score	E-value
Q662Q2_BORGA	Putative uncharacterized protein	24	1.4
Q660D1_BORGA	Putative uncharacterized protein	23	3.1
Q663A7_BORGA	Tryptophanyl-tRNA synthetase	22	5.3
Q661J8_BORGA	Sensory transduction histidine/ kinase response regulator	22	5.3
Q661F1_BORGA	Putative uncharacterized protein	22	6.9

**Table 3.1. BLAST search of 1BXW from *E. coli* against the *B. garinii* proteome.** The score represents the quality of the match between two sequences; the higher the score, the better the match. The (Expect) E-value is the number of matches that can be expected to occur at random. The lower the E-value, the more significant the result; a value between 0.1 and 10 indicates that match is uncertain and a value of 10 or more indicates that there is likely no biological significance (Altschul *et al.*, 1997). The scores of the potential matches are very low, which is to be expected due to the high sequence variability of TM  $\alpha$ -barrel proteins. The E-values are also high, a positive control search of 1BXW against the *E. coli* K 12 proteome returned 1BXW itself as the top hit with a score of 362 and an E-value of e-102. Of the proteins identified, three have known functions and the remainder have molecular weights more than twice the size of an eight stranded TM  $\alpha$ -barrel protein.

No matches were found due to BLAST searches not being sensitive enough to detect the substantial sequence divergence characteristic of OmpA-like proteins (Schulz, 2000). Hidden Markov models (HMMs) were used as an alternative (Bernsel *et al.*, 2008), the first step of which was building a HMM profile of known OmpA-like proteins.

### 3.1 Classification of known Omps

A list of 167 known OmpA-like transmembrane domains from multiple *Enterobacteriaceae* (ProDom family PD004606) (Servant *et al.*, 2002) was used to provide a broad reference for searching the *Borrelial* genomes for potential OmpA-like proteins. To expedite the search the list was trimmed to remove any proteins with a sequence more than 40 % identical to any other protein in the list using the Decrease Redundancy program developed by Cédric Notredame (unpublished) hosted on ExPASy (Gasteiger *et al.*, 2005). A second list was produced with a 90 % identity limit.



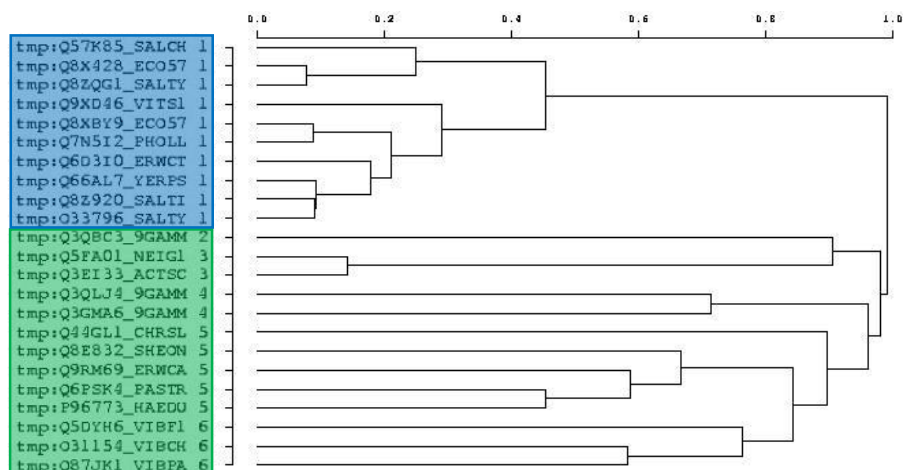
The Decrease Redundancy program works by aligning and comparing all sequences pairwise, before measuring the percentage ID between the first two sequences. When a pair is found with a percentage identity greater than the specified cut-off, one of the sequences is removed and the accession number placed in the information line along with the accession number of the sequence kept upon which the program then moves on to the next sequence in line. This continues until all of the sequences in the file have been compared and any with a percentage identity greater than the specified have been removed. The results of both reductions are included in the appendix 8.1, all similar proteins are displayed as in Figure 3.1.

```
>Q87JK1_VIBPA      Q8D6H3_VIBVU      Q7MDC9_VIBVY
Q87GX9_VIBPA

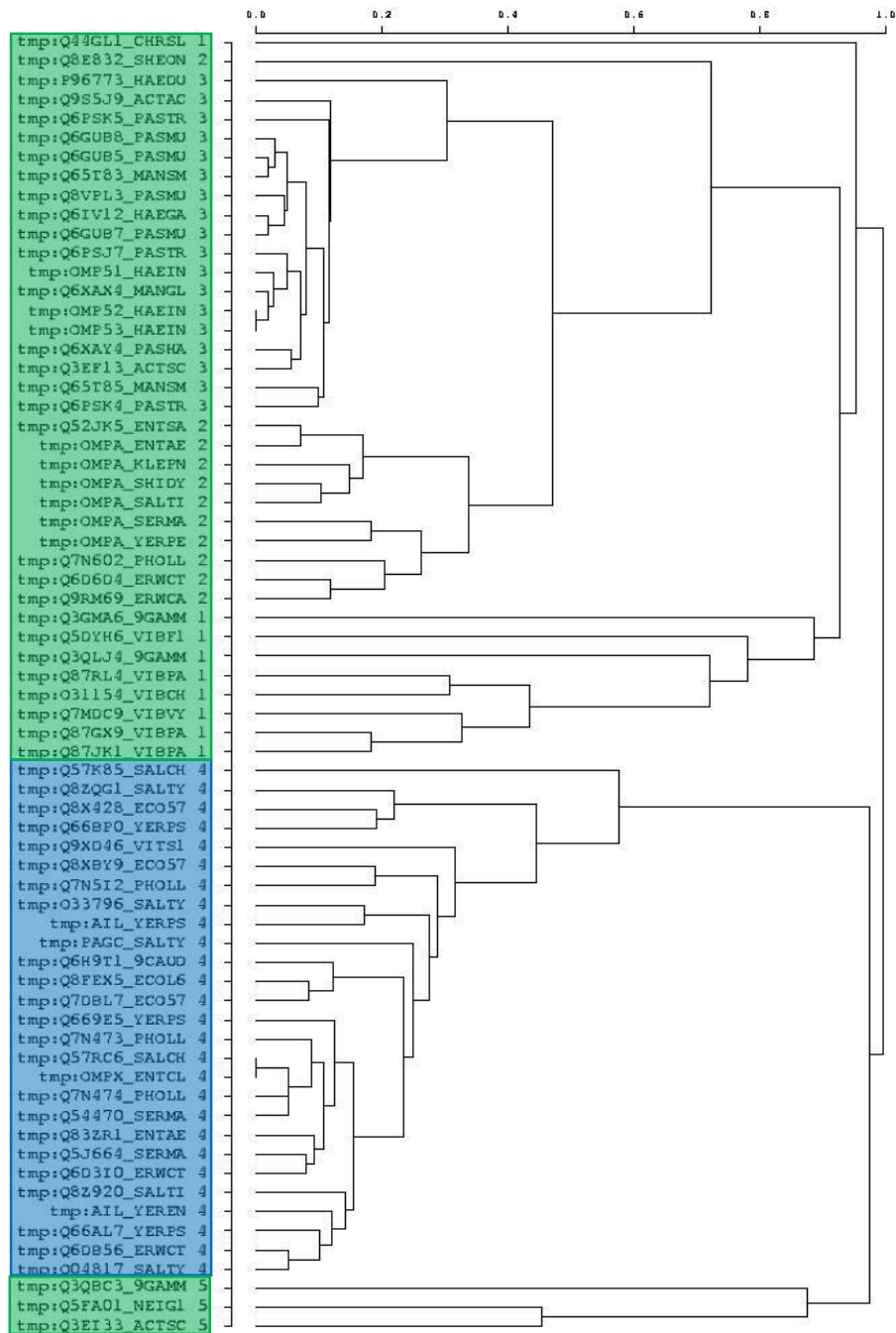
YIGGKVGYNLNDACYLNEPCDDDSFAAGMHIGYNFNEYVAAEYGVVDYLGDFLTANFKKPGLN
TVDGDLWALTLPKFNLPDTSWNLFAKVGAAAYMMAGDEKDFVPTGSLGAEYKINYNWSLRA
EYQRYQDMSDDIVDDMDSDFFGIGFNYKFG
```

**Figure 3.1. An example of a redundancy reduced protein search.** Proteins more than 40 % identical to a protein already observed by the program have the accession numbers included above the sequence of this initial protein, next to its accession number.

The lists of known OmpA-like proteins produced from both the 40 % and 90 % redundancy reductions were submitted to the JACOP (Just Another Classification Of Proteins) server (Sperisen and Pagni, 2005) for classification (Figure 3.2 & Figure 3.3). Through using random sampling of short sub-sequences and comparing them to the initial input of proteins, scores based on similarity are assigned, which are used to classify the proteins and organise them according to similarity in the form of New Hampshire trees (Sperisen and Pagni, 2005). Each protein present in the New Hampshire trees was searched in the UniProtKB database to identify the class of OM protein - OmpA/W/X.



**Figure 3.2. NewHampshire Tree produced using a list of known TM  $\alpha$ -barrel proteins reduced by 40 %.** Image generated by JACOP from a list of known TM  $\alpha$ -barrel proteins with a redundancy reduction of 40 %. Highlighted are the two main classes of OM protein in this tree. Proteins within the green box have been identified as OmpA and within the blue box OmpX. Interestingly no OmpWs were identified. The numbers to the right of the accession numbers signify the grouping assigned by JACOP.



**Figure 3.3. NewHampshire Tree produced using a list of known TM -barrel proteins reduced by 90 %.** Image generated by JACOP from a list of known TM -barrel proteins with a redundancy reduction of 90 %. The classes of OM protein are highlighted according, and again there were no OmpW proteins identified. The groups of OmpA and OmpX are both rather larger in the 90 % reduced NewHampshire tree compared to the 40 % due to a higher similarity cut-off. Though there are 5 numbered groups, it was shown by looking up the accession numbers in the RSCB PDB that groups 1-3 are all OmpA from various bacteria; the numbered groups signify that some proteins are more closely related than others.

Each of the redundancy reduced protein lists were aligned automatically using ClustalW2 (Goujon *et al.*, 2010, Larkin *et al.*, 2007). The 40 % reduced sequence alignment is included (Figure 3.4) with conserved residues highlighted.

Q8XBY9_ECO57#13#177	.....	..SAGYAQSK	VQDFKN....	.IKG.VNLQY	RYEWDSP.VS
Q7N5I2_PHOLL#43#182	.....	.....	.....	.LNG.ATLSY	RYELDDQ.WG
Q57K85_SALCH#72#178	.....	.....	.....	...G.S...Q	RYSDSS.NG
Q6D3I0_ERWCT#17#171	GAGSAFAGQS	TVTAGYAQGD	AQGVQN....	KVKG.FNLKY	RYEQDNNPLG
Q9XD46_VITS1#32#191	.....	TVSIGYSHGK	ISGAD....	KLNG.VTAKY	NYQFDQQPWG
Q66AL7_YERPS#27#179	.....ASN	TVSFGYAQST	LKIDGKIGK	DNKG.FNLKY	.RHELDSDLG
Q8X428_ECO57#7#108	.....	SVMAG...PS	VRVN.....	.....	.....
O33796_SALTY#23#153	.....AGYKN	TVSIGYAYTD	LSGWLSGNAN	...G.ANIKY	NWEDLDSGFG
Q8ZQG1_SALTY#92#172	.....	.....	.....	.....	.....
Q8Z920_SALTI#29#174	.....	..SFGYAQTH	LSSLKNSDSK	DLRG.FNFKY	RYEFNET.WG
Q5FA01_NEIG1#20#175	....EGASGF	YVQADAAHAK	ASSS..LGS.	.AKG.FSPRI	SAGYRINDLR
Q3EI33_ACTSC#24#191	.....	YVEGDAGYSK	LRTSGGLSSK	ISDGSFSPSV	AVGYKVNDWR
Q9RM69_ERWCA#29#207	..YTGGKLG	VQFHDGTGYG	NGYTDVNNNP	IKSKLGAGAF	VGYQANPYLG
Q8E832_SHEON#33#191	..YAGAGLG.	.....QGHYS	NGSNPQSYDS	VRDRFAGSVY	LGQVNPYLA
Q5DYH6_VIBF1#21#185	..YLGGQVG.	.....VSHFL	GACSSNAIEC	KNYVTGGGLY	GGYQFNSWFA
O31154_VIBCH#24#173	..YVGGKVG.	.....KSWL	DDACLAGQSC	EDDDQVVGAF	LGQYQANKWLS
Q87JK1_VIBPA#26#179	..YIGGKVG.	.....YNSL	NDACYLNEPC	DDDSFAAGMH	IGYNFNEYVA
Q6PSK4_PASTR#28#221	..YVGAKAGW	ASFHDGISQI	DHKNNGKGYI	NKNSVTYGAF	VGYQIIDNLA
P96773_HAEDU#126#257	.....	.....	.....	.....	.....
Q3QLJ4_9GAMM#38#150	.....	.....	...ASGDGEG	SSDGVGFGAY	GCYNFNEWFG
Q3GMA6_9GAMM#36#179	.....	....YVGAK	VGQFDLDVDG	ADDPTAYGVY	AGYNFDPNFG
Q3QBC3_9GAMM#28#162	.....	....AKGGYQ	WASDDSYKHS	NPKGAIFGVS	SGLQFSPAWS
Q44GL1_CHRSL#62#200	.....	.....	.....DV	DDDDNTFKGF	VGYNFNRYFA
Q8XBY9_ECO57#13#177	VVGSFSYMKG	DWADSHRDEA	DDFYRHQADI	KYYSFLAGPA	YRLNDY..ISF
Q7N5I2_PHOLL#43#182	LLSSFTFAKG	.....DEK	ETKYTGEDL	KYYSVMIGPT	YRLNDY..ISL
Q57K85_SALCH#72#178	RVTT.....	.....	.....	RYYSLLAGPS	WKINNQ..LSL
Q6D3I0_ERWCT#17#171	VISSFTYLE.	..KN..GSDD	GFYS...KG	QYMGFTAGPA	YRLNDW..ASL
Q9XD46_VITS1#32#191	VMTNLTYMGG	KQRNNTSSN	QLYEN.DVDV	DYYSAGVGPS	YRINPN.VNV
Q66AL7_YERPS#27#179	IVASFTHTKQ	NYGMPGSDSG	K.....RKV	EYYSLMVGPS	WRFNEF.VSA
Q8X428_ECO57#7#108	..EWFS....	AYAMAG....	.....V	AYS.....	.RVSTF.XGD
O33796_SALTY#23#153	AMGSVTTYTSA	DVNNGYKVG	.....DA	DYTSLLVGPS	YRFNDY..LNA
Q8ZQG1_SALTY#92#172	.....	.....	.....	.....GPS	YRFNEY..LNA
Q8Z920_SALTI#29#174	MLGSFTATR	EMENYTWKEG	KLHKNGSDSV	DYGSIMFGPT	YRFNDY..VSL
Q5FA01_NEIG1#20#175	FAVDYTRYKN	.....YKQ	APST....DF	KLYSIGASVI	YDFDTQ.SP
Q3EI33_ACTSC#24#191	FALDYTTYGK	GDEGFGDLGG	SGEN....EV	KAYGFGLAAY	YDFDLG.TSL
Q9RM69_ERWCA#29#207	FEMGYDWLGR	M.KYAGSTAN	PADS.ASLKA	QGIQLAAKLS	YPVLPD.LDV
Q8E832_SHEON#33#191	PELSYQFLG.	....SAYANY	EQQG.ISGDF	QQVVLAAARFG	YPLTTS..LYP
Q5DYH6_VIBF1#21#185	LEGSWHDYG.	....SPKTFY	GVGDGYYSNA	TGVDLSVKLS	LPVTDN.LDL
O31154_VIBCH#24#173	LEAGYDYLG.	..KFTA....	AGLN..DEKV	QAVTLAPKLS	IPLTEG..IAL
Q87JK1_VIBPA#26#179	AEYGVGYLG.	..DFTANFKK	PGLNTVDGDL	WALTLAPKFN	LPLTDS.WNL
Q6PSK4_PASTR#28#221	AEVGYEYFGR	VRGLEQTKAG	GSKKTFRHS	HGTTIALKGN	YEVISG.LDT
P96773_HAEDU#126#257	.....	.....KG	DLNRQFKHTA	HGANLSLKPS	YEILPN.LDV
Q3QLJ4_9GAMM#38#150	LEANLFASG.	.....D	LGEDGVDVGA	GALTFTPKFT	YHFNDT.FSA
Q3GMA6_9GAMM#36#179	MEAEFVGSD.	.....D	ADYRGGDIDA	KTYGAYGTYR	YAFNPNTGLYA
Q3QBC3_9GAMM#28#162	WDVGYQYHD.	.....DLK	ADATSVNVKA	GLIESALRYD	WYLQDN.LSV
Q44GL1_CHRSL#62#200	TEAFYSDLG.	....RVKLLG	NGTANTDLES	EAYGVSLVGK	LPITQW.FEL
Q8XBY9_ECO57#13#177	YGLVGISHTK	AKGD.....	.YEWNSVGA	DESDGYLSES	VSKKSTD.FAY
Q7N5I2_PHOLL#43#182	YQQLGLSRIN	DK.....	.....ST.A	HYSGGYIEKE	STSKNT.LGW
Q57K85_SALCH#72#178	YSQVGPVLLH	QR.....	.....DHGI	NESDSKVG..	.....YGY
Q6D3I0_ERWCT#17#171	YGVVGFSS..	.HGK.....	.VTSNNTNGQ	DNASNDGYG.	.....FTY
Q9XD46_VITS1#32#191	YGVVGVAKAD	TKGT.....	.QKQNRGTGRI	YNIDNIDTN.	.....VMY
Q66AL7_YERPS#27#179	YALIGAT...	.QK.....	.STHTKPRMV	SNTVSKTS..	.....MGY
Q8X428_ECO57#7#108	YLRVTDN...	.KGK.....	.THDVLGTGSD	DGRHSNTS..	.....LAW
O33796_SALTY#23#153	YVMIGAA...	.NGH.....	.IKDNW....	GNSDNKTA..	.....FAY
Q8ZQG1_SALTY#92#172	YVMAGLG...	.HGH.....	.IDDKR....	DNSGKKTG..	.....FAY
Q8Z920_SALTI#29#174	YGNAGIA...	.TMK.....	.FN.....	.KHSKEES..	.....FAY
Q5FA01_NEIG1#20#175	KPYFGARLSL	N..R.....	.ASAHLG.GS	D....SFSKT	SAG....LGV
Q3EI33_ACTSC#24#191	TPYLGARISA	NHIK.....	.VNSDFNNAS	DYFHFSNSNT	EFG....YGA
Q9RM69_ERWCA#29#207	YTRLGGMVWR	VDTH.....	.ADRSGNH.L	NNDDTGVS..	.....PLA
Q8E832_SHEON#33#191	YVKVGAGWF	GDSE.....	.GLRSG....	..SERGFS..	.....PIV
Q5DYH6_VIBF1#21#185	YAKGGAAYNY	LSVS.....	.GNDNVHLGM	FESDSSID..	.....DIW
O31154_VIBCH#24#173	YGVVGGAYVD	YG.....	.....	..SKDDYS..	.....YLG
Q87JK1_VIBPA#26#179	FAKVGAAAYM	AG.....	.....	..DEKDFV..	.....PTG
Q6PSK4_PASTR#28#221	YAKAGIALVN	NSYK.....	TVNVDTKVPT	KTSRFQSS..	.....LIL
P96773_HAEDU#126#257	YGVVGMGLVR	NDYKFILDLO	SKDQHVKDET	KLHTLKPS..	.....LLL
Q3QLJ4_9GAMM#38#150	FAKVGVASMA	VVVD.....	.....YGP	DADYTGWG..	.....LTW

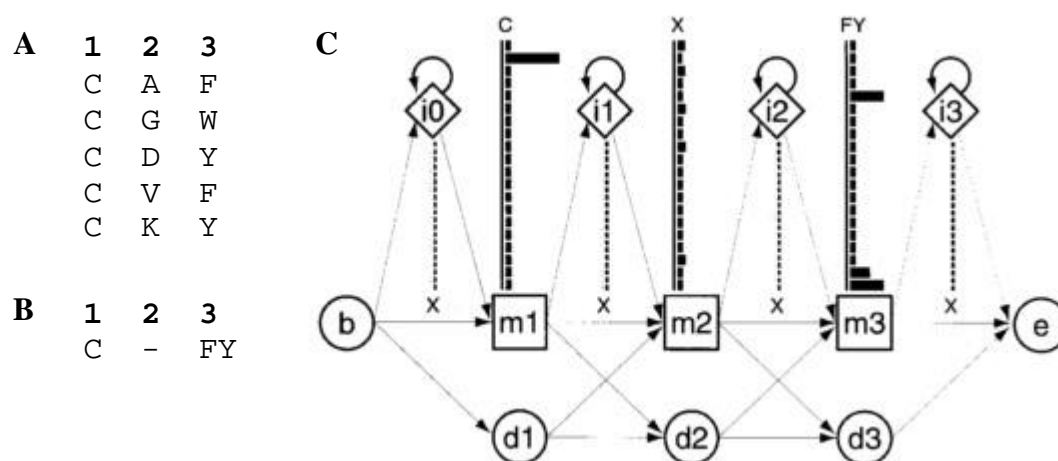
Q3GMA6_9GAMM#36#179	KGKLGVARTE	IEATG.....	...TSPAGFY	NGSNKDTS..	.....LAG
Q3QBC3_9GAMM#28#162	YGRLGVAYWD	MEKT.....	.....HLSSD	KLDATGLS..	.....PLG
Q44GL1_CHRSL#62#200	FAKAGMAKWE	TDIDG.....	..NLGGASTD	LEDNDGVD..	.....PVY
Q8XBY9_ECO57#13#177	AAGVIINPWG	NMSVNVGYEG	TKADIYG...	.....	..KHSVNGFT
Q7N5I2_PHOLL#43#182	GAGFIINPTT	NTSITAGYEG	SRFSIKDGNE	K.....	..DHLSTNGFN
Q57K85_SALCH#72#178	SAGVAYTPVS	NVAITLGYES	ADFDATHNS.	.....	..GSLNSNGFN
Q6D3I0_ERWCT#17#171	GAGVQFNPIQ	DVALDVGYES	SRIRS.....	.....	...VDVGAWA
Q9XD46_VITS1#32#191	GAGVQYNPAP	NWSVLDVYES	SRVNDGYDK.	.....	...RSMNAFN
Q66AL7_YERPS#27#179	GAGLQFNPKV	HVAIDTAYEY	AKIEDVKIG.	.....	...TWI....
Q8X428_ECO57#7#108	GAGVQFNPEX	SVAIDIAYES	SGSGDWRTD.	.....	...GFI....
O33796_SALTY#23#153	GAGIQLNPVE	NIAVNASYEH	TSF.....	.....	.....
Q8ZQG1_SALTY#92#172	GAGVQINPVE	NIAINASYEY	SRFSAYDSK.	.....	...VNAGTWV
Q8Z920_SALTI#29#174	GAGVIFNPVK	SISIDASWEA	SRFFAVDTN.	.....	...TFG....
Q5FA01_NEIG1#20#175	LAGVSYAVTP	NVDLDAGYRY	NYVGKVNNVK	.....	..NVRSGELS
Q3EI33_ACTSC#24#191	IAGVSYNFAP	QWDLNVAAEY	NRLGEVRDVK	.....	...VNQYG.AK
Q9RM69_ERWCA#29#207	AIGIEYAIID	NWATRVDYQW	VSNIGDA...	.....	..GTG GARPDMMLMS
Q8E832_SHEON#33#191	AAGVEYAFTP	RLSGRLEHQY	TDSLGA...	.....	..DSI GYT.DHHLTT
Q5DYH6_VIBF1#21#185	EIGAEYALAP	NWSLRFGTST	IDGIG.....	.....	..NAK TGKSDLYFTS
O31154_VIBCH#24#173	AAGLEFNTNH	NVTMRLEYQN	LTDIN.....	.....	..NDI VR.ARAETAT
Q87JK1_VIBPA#26#179	SLGAEYKINY	NWSLRAEYQR	YQDMS.....	.....	..DDI VDDMDSDFFG
Q6PSK4_PASTR#28#221	GAGVEYAITP	SLGARIEYQW	LNNAGKASYA	TLRRMGVEGS	DYRPDISSVS
P96773_HAEDU#126#257	GTGVEYAITP	ELALRTEYQY	LNKAGNLYKA	AKYKNISTSL	AYAPDIHSVS
Q3QLJ4_9GAMM#38#150	GVGVNAALTE	KLNLRLSY..	.....	.....	.....
Q3GMA6_9GAMM#36#179	GVGLGYSVNP	NFSVLAHEYDK	LG.....	.....	...SDADLMT
Q3QBC3_9GAMM#28#162	EVGVNYNFTP	NVRLSAGYQY	IDSIGESN..	.....	.....
Q44GL1_CHRSL#62#200	GAGAQNFKKP	.FLVRAEYER	YDFDS.....	.....	..DYQIDSFT
Q8XBY9_ECO57#13#177	VGVGYRF...	..			
Q7N5I2_PHOLL#43#182	ITVGYRF...	..			
Q57K85_SALCH#72#178	LGVGYRF...	..			
Q6D3I0_ERWCT#17#171	VGVGYRF...	..			
Q9XD46_VITS1#32#191	VGVGYRF...	..			
Q66AL7_YERPS#27#179	VGVGYRF...	..			
Q8X428_ECO57#7#108	VGVGYKF...	..			
O33796_SALTY#23#153	.....	..			
Q8ZQG1_SALTY#92#172	LGVGYSF...	..			
Q8Z920_SALTI#29#174	VSVGYRF...	..			
Q5FA01_NEIG1#20#175	AGVRVKF...	..			
Q3EI33_ACTSC#24#191	VGVRVTF...	..			
Q9RM69_ERWCA#29#207	VGLSYRFGQD	D.			
Q8E832_SHEON#33#191	LGLSWRFG...	..			
Q5DYH6_VIBF1#21#185	LGLTYKF...	..			
O31154_VIBCH#24#173	LGLIAYKFGGS	E.			
Q87JK1_VIBPA#26#179	IGFNYKFG...	..			
Q6PSK4_PASTR#28#221	AGLTYRFGQG	AA			
P96773_HAEDU#126#257	VGLSYRFGQG	..			
Q3QLJ4_9GAMM#38#150	.....	..			
Q3GMA6_9GAMM#36#179	VGAQLKF...	..			
Q3QBC3_9GAMM#28#162	.....	..			
Q44GL1_CHRSL#62#200	ASVGVQF...	..			

**Figure 3.4. Sequence alignment of <40 % identical known TM  $\alpha$ -barrel proteins.** This sequence alignment was produced using ClustalW2 (Goujon *et al.*, 2010, Larkin *et al.*, 2007) from the list of proteins with a redundancy reduction of 40 %. There are a relatively large number of conserved residues (turquoise), which appear to be grouped together (~10 - 20 residues), and large stretches (> 20 residues) of highly variable regions. The conserved regions correspond to TM  $\alpha$ -strands, with the variable stretches being the loop regions. Highlighted in yellow (positive) and red (negative) are conserved charged residues present within a  $\alpha$ -strand.

### 3.2 Profile hidden Markov model searches

Hidden Markov model (HMM) profile searching can be used either in place of or in conjunction with traditional searches such as BLAST to identify homologous proteins. HMM profile searches were initially applied to speech pattern recognition (Eddy, 1998), but have since been developed to compare and contrast protein sequences from a multiple sequence alignment (MSA) to either single sequences or other profile HMMs (Söding, 2005). MSAs of a protein family highlight the most conserved residues and regions that are more likely to contain mutations, insertions or deletions (Eddy, 1995, Eddy, 1998), therefore comparing HMM profiles of MSAs are more likely to result in identification of protein homologues with large sequence variability than BLAST searches, which compare one sequence directly with another, affording equal weighting to each residue.

Hidden Markov modelling involves the production of a statistical model of protein structure consensus (a profile) (Eddy, 1995, Eddy, 1996, Eddy, 1998). A HMM profile is a full probabilistic primary model (Eddy, 2004) of the structure consisting of position-specific residue scores modified depending on match, insertion and deletion states along each column in the MSA (Eddy, 1996). The match state models the distribution of residues present in each column, while the insert and delete states correspond to the position-specific probabilities for transition from one state to another (Eddy, 1996); in other words the insertion or deletion of one or multiple residues between that column and the next (Figure 3.5). Profile HMMs therefore account for the conservation of residues along with the statistical potential for variability or insertions/deletions (Eddy, 1996, Eddy, 1998). A HMM profile built from homologous MSAs contains more information about the family of sequences than could be obtained from a solitary sequence. Also included in the profile is the individual probability of each residue being present at each insertion state (Eddy, 1998, Söding, 2005).



**Figure 3.5. A small HMM profile.** **A)** a multiple sequence alignment (MSA), **B)** Consensus sequence derived from **A**, **C)** a representation of the HMM. The HMM (**C**) is representative of the small MSA (**A**). The MSA consists of five sequences, showing only the first three residues. The HMM has a beginning and an end ('b' and 'e' respectively). The distribution of residues in each column in the MSA is modelled by a match state (m1 to m3) in the HMM. The match state has 20 residue emission probabilities (the black bars above the match state), as does each insert state (i0 to i3). The insertion state takes into account the insertion of one or multiple residues between columns. The delete states have no inherent emission probabilities and allow for deletion of the next consensus residue 'expected' from the MSA. Arrows represent state transition probabilities. Image adapted from (Eddy, 1998, Eddy, 1995).

Proteome searching using HMM profiles, like BLAST searches produces an output with two important values; score and E-value. The score is a measure of the match of a given protein sequence to the HMM profile, the higher the score the better the match (Eddy, 2010). The E-value (expectation value) is a measure of the statistical significance of the calculated score of the match to the sequence (Eddy, 2003, Eddy, 2010). The lower the E-value the greater the significance of the match. The E-value is related to the number of sequences in the database being searched and can be interpreted as the expected number of false positives expected to have that score within the search database (assuming that the sequences are random and non-homologous) (Eddy, 2010), therefore identical sequences in databases of different size would have different E-values.

HMMER v2.3.2 was used to build and calibrate HMM profiles of each of the MSAs created previously (Section 3.1). These profiles were visualised using the online server LogoMat-M (Schuster-Boeckler *et al.*, 2004) to provide an visual representation of conserved residues (Figure 3.6).







The HMM profiles were then used to search the proteomes of *B. afzelii*, *B. burgdorferi* s.s. and *B. garinii* for OmpA-like protein homologues, again using HMMER v2.3.2. All HMM searching was carried out using the eridani cluster at the High Performance Computing (HPC) at the University of Huddersfield. Results obtained from searching HMMs of known OmpA-like proteins from each proteome are below (Table 3.2). Proteomes were obtained from the UniProtKB database (Magrane and Consortium, 2011) and are composed of the following numbers of proteins: *B. afzelii* (PKo); 1,695 proteins, *B. burgdorferi* s.s. (ZS7); 1,223 proteins, *B. garinii* (PBi); 1,262 proteins.

Proteome	RR (%)	Reference numbers	Ordered Locus name	Description	Score	E-Value	Seq length (a-a)	Mass (Da)	
Bg	90	Q08PD7	Q08PD7_BORAP	BAPKO_0026	Putative uncharacterised protein	-71.2	0.3	211	23843
		Q08M57	Q08M57_BORAP	BAPKO_0847	Putative uncharacterised protein	-80.2	1.8	1467	169235
		Q08SNC3	Q08SNC3_BORAP	BAPKO_0339	Basic membrane protein C	-83.4	3.2	353	39572
		Q08SNQ2	Q08SNQ2_BORAP	BAPKO_0269	Putative uncharacterised protein	-83.8	3.5	722	85085
		Q08SLQ6	Q08SLQ6_BORAP	BAPKO_2525	Putative uncharacterised protein	-86.8	6.1	391	45913
		Q08SNQ9	Q08SNQ9_BORAP	BAPKO_0262	Conserved hypothetical integra	-88.0	7.7	767	86754
		Q08SM13	Q08SM13_BORAP	BAPKO_0891	Putative uncharacterised protein	-88.4	8.3	1133	132442
		Q08SNA2	Q08SNA2_BORAP	BAPKO_0422	Putative uncharacterised protein	-88.4	8.4	201	22232
		Q08SLH2	Q08SLH2_BORAP	BAPKO_3514	Putative uncharacterised protein	-89.0	9.4	198	23548
		Q08SM34	Q08SM34_BORAP	BAPKO_0870	UDP-N-acetylmuramate-L-alanine	-89.0	9.4	468	53553
		Q08SLB8	Q08SLB8_BORAP	BAPKO_4013	Putative uncharacterised protein	-60.1	2.3	434	48602
		Q08SNC3	Q08SNC3_BORAP	BAPKO_0399	Basic membrane protein C	-61.1	2.7	353	39572
		Q08SNS5	Q08SNS5_BORAP	BAPKO_0245	Putative uncharacterised protein	-63.8	4.6	668	75489
		Q08SKY4	Q08SKY4_BORAP	BAPKO_6013	Putative uncharacterised protein	-64.6	5.2	251	24033
Q08SND5	Q08SND5_BORAP	BAPKO_0387	Putative uncharacterised protein	-65.4	6.1	220	25487		
Bt	90	Q08SM57	Q08SM57_BORAP	BAPKO_0847	Putative uncharacterised protein	-67.4	8.9	1467	169235
		O51734	O51734_BORBU	BB_0794	Putative uncharacterised protein	-74.0	0.9	1465	168804
		O51058	Y027_BORBU	BB_0027	Uncharacterised protein	-75.0	1.1	212	23865
		O51274	O51274_BORBU	BB_0259	Putative uncharacterised protein	-81.8	3.6	737	86337
		O51213	O51213_BORBU	BB_0195	Cell division control protein	-82.1	3.8	379	44124
		Q59181	PGK_BORBU	BB_0056	Phosphoglycerate kinase	-82.7	4.2	393	42346
		O51493	O51493_BORBU	BB_0543	Putative uncharacterised protein	-82.7	4.2	218	24388
		O51486	O51486_BORBU	BB_0536	Zinc protease, putative	-83.2	4.6	933	108457
		Q44881	Q44881_BORBU	BB_0603	Membrane-associated protein p6	-84.1	5.4	618	68172
		O51187	O51187_BORBU	BB_0165	Putative uncharacterised protein	-85.9	7.4	614	71411
		Q45011	BMPB_BORBU	BB_0382	Basic membrane protein B	-86.6	8.4	341	37549
		O51137	O51137_BORBU	BB_0110	Putative uncharacterised protein	-87.3	9.5	454	53411
		O51252	O51252_BORBU	BB_0236	Putative uncharacterised protein	-83.3	1.4	668	75350
		O50780	O50780_BORBU	BB_225	Putative uncharacterised protein	-61.5	5.3	350	40712
Q45010	BPMA_BORBU	BB_0383	Basic membrane protein A	-63.7	7.6	339	36968		

<i>Bg</i>	90	Q662Y5	Q662Y5_BORGA	BG0027	Purative uncharacterised protein	-75.8	1.3	212	23989
		Q5XYQ8	Q5XYQ8_BORGA	BGP223	Purative uncharacterised protein	-76.9	1.5	1012	96689
		Q65ZX9	Q65ZX9_BORGA	BG0820	Purative uncharacterised protein	-77.7	1.7	1466	168955
		Q662A6	Q662A6_BORGA	BG0262	Purative uncharacterised protein	-81.8	3.4	715	83813
		Q662Q2	Q662Q2_BORGA	BG0111	Purative uncharacterised protein	-81.9	3.5	454	53409
		O31362	BMPB_BORGA	BG0381	Basic membrane protein B	-84.5	5.4	341	37211
		Q661P1	Q661P1_BORGA	BG0377	Purative uncharacterised protein	-84.5	5.4	219	25524
		Q662B3	Q662B3_BORGA	BG0255	Conserved hypothetical integral	-85.4	6.3	766	86895
		Q662H0	Q662H0_BORGA	BG0193	Purative uncharacterised protein	-85.7	6.6	379	44249
		Q660W1	Q660W1_BORGA	BG0573	Purative uncharacterised protein	-86.1	7.1	172	18797
	40	Q662J4	Q662J4_BORGA	BG0169	Purative uncharacterised protein	-86.8	8.0	683	78355
		Q662V6	Q662V6_BORGA	BG0057	Purative uncharacterised protein	-87.1	8.4	656	76550
		Q662M0	Q662M0_BORGA	BG0143	Purative uncharacterised protein	-60.5	3.9	411	47121
		O31357	BMPA2_BORGA	BMPA2	Basic membrane protein A2	-61.5	4.6	337	36878
		Q5XYQ8	Q5XYQ8_BORGA	BGP223	Purative uncharacterised protein	-62.2	5.2	1012	96689
		Q662J4	Q662J4_BORGA	BG0169	Purative uncharacterised protein	-64.6	7.6	683	78355
		Q661N6	BMPA1_BORGA	BMPA1	Basic membrane protein A1	-65.6	8.9	337	36862
		Q662C8	Q662C8_BORGA	BG0239	Purative uncharacterised protein	-66.0	9.4	668	75255

**Table 3.2. HMM profile search results of *Borrelial* proteomes.** Highlighted in green are the putative proteins with approximately the expected sequence length and molecular weight of OmpA-like proteins. Reference numbers are taken from the online database BioCyc. All proteins highlighted in the table have a negative score, which is an indication of a sequence that is a poor match to the profile. Each of the highlighted proteins (with the exception of BAPKO\_0026) has an E-value of above 1, which would suggest that the result could be a false positive, however as stated previously the E-value is subject to the size of database being searched (in this case the proteomes being searched each contain over 1,000 proteins), hence a score of 1 would suggest that out of 1,000 sequences, 1 could match that score by chance. The low scores and relatively good E-values (<10) indicate that the sequences identified are homologous to the family modelled by the HMM, but not within the family itself, in this case the proteins would be OmpA-like rather than an OmpA, W or X (Eddy, 2003).

The UniProtKB database was used to assign putative functions to all proteins identified by HMM searches (Table 3.2). Several proteins are unlikely to be OM-spanning proteins based on homology to proteins of known function and/or molecular weights considered to be larger than would be predicted for an eight-stranded membrane spanning  $\alpha$ -barrel protein. However eight proteins have molecular weights comparable to *E. coli* OmpW/X and are potential candidates to be OmpA-like. BLAST searches of these candidates showed that BAPKO\_6013 is homologous to a VlsE protein and BAPKO\_3514 is homologous to a plasmid-encoded Type I restriction enzyme R and so both are unlikely to be OmpA-like proteins. The other proteins had no identifiable homologues and are potential candidates to be *Borrelial* OmpA-like proteins.

Orthologues in each of the three main *Borrelia* species were identified using a multi-genome browser (BioCyc) and the accession numbers displayed below (Table 3.3).

<i>B. afzelii</i>	<i>B. burgdorferi</i> s.s	<i>B. garinii</i>
BAPKO_0026	BB_0027	BG0027
BAPKO_0387	BB_0378	BG0377
BAPKO_0422	BB_0405	BG0407
BAPKO_0571	BB_0543	BG0553
BAPKO_0423	BB_0406	BG0408

**Table 3.3. Summary of potential OmpA-like proteins identified in *Borrelia*.** Putative proteins identified through HMM searches and manual inspection. BAPKO\_3514 and its homologues are not included as this was found to be a type I restriction enzyme in the UniProtKB database (Magrane and Consortium, 2011). Proteins are identified by the locus name. Orthologues retrieved from the BioCyc database from proteomes of *B. afzelii* (PKo), *B. burgdorferi* s.s. (Zs7) and *B. garinii* (PBi) (Caspi *et al.*, 2010). BAPKO\_0423, BB\_0406 and BG0408 are paralogues of BAPKO\_0422, BB\_0405 and BG0407, likely the result of a gene duplication event. These proteins were identified using the genome browser on the BioCyc webserver (Caspi *et al.*, 2010).

Because of the remote homology and high sequence divergence of OmpA proteins (Schulz, 2000) a high number of false positives is expected, therefore fold prediction algorithms (FFAS03) and other prediction tools were used to corroborate the results (Section 3.3).

### 3.3 Topology predictions

Topology predictions were carried out on putative OM spanning  $\alpha$ -barrel proteins identified by HMM searches and genome browsing (Table 3.3) using the online Function & Fold Assignment (FFAS03) server (Jaroszewski *et al.*, 2005) and transmembrane  $\alpha$ -barrel

prediction server PRED-TMBB (Bagos *et al.*, 2004), along with analysis of the signal peptide.

### 3.3.1 FFAS03 assessment

FFAS is a profile-profile and fold recognition algorithm that is capable of identifying conserved residues and applying this to the detection of remote homologues. A target sequence (in this case potential TM  $\alpha$ -barrel protein) is used to create a profile which is used to search several protein and domain databases (in this case PDB0113) (Jaroszewski *et al.*, 2005). Each hit identified in Table 3.3 was searched using FFAS03 to identify remote homologues. The lower the score reported, the greater the confidence in the result; scores of lower than -9.5 contain have less than a 3 % chance to be a false positive result (Jaroszewski *et al.*, 2005). Included (Table 3.4) is an example of results from an FFAS03 search of BG0407.

#	Score	Template	% ID
1	<b>-9.970</b>	2x27_X OUTER MEMBRANE PROTEIN OPRG	10
2	<b>-9.820</b>	2k0l_A Outer membrane protein A	13
3	-9.250	2f1t_A Outer membrane protein W	9
4	-8.920	1p4t_A outer membrane protein NspA	11
5	-8.010	2lhf_A Outer membrane protein H1	14
6	-7.660	2ge4_A Outer membrane protein A	13
7	-7.560	2jmm_A Outer membrane protein A	17
8	-6.880	1bxw_A PROTEIN (OUTER MEMBRANE PROTEIN A)	13
9	-6.460	2o6l_A UDP-glucuronosyltransferase 2B7	11
10	-6.390	1orm_A Outer membrane protein X	11

**Table 3.4. FFAS03 results from BG0407.** A score more negative than -9.5 signifies a good result (less than 3 % chance of being a false positive). The top hits all have very low % sequence identity to BG0407. Nine of the top ten hits are OM proteins, and the top two hits have a score below -9.5, showing with high probability that BG0407 is highly likely to be an OM protein. Of interest are the third and eighth hits 2F1T and 1BXW; OmpW and OmpA from *E. coli* respectively.

FFAS03 analysis was carried out on all proteins in Table 3.3 and the results summarised in Table 3.5.

Locus	Best score	Locus	Best score	Locus	Best score
BAPKO_0026	<b>-36.20</b>	BB_0027	<b>-47.90</b>	BG0027	<b>-40.60</b>
BAPKO_0387	-7.71	BB_0378	-8.72	BG0377	-8.93
BAPKO_0422	-8.94	BB_0405	<b>-10.10</b>	BG0407	<b>-9.97</b>
BAPKO_0423	<b>-9.54</b>	BB_0406	<b>-11.80</b>	BG0408	<b>-12.30</b>
BAPKO_0571	-7.060	BB_0543	-7.33	BG0553	<b>-9.72</b>

**Table 3.5. Summary of FFAS03 results.** Unless stated the top hit was either an OmpA or OmpW protein and scores more significant than the -9.5 cut-off are highlighted in bold. BAPKO\_0026 and homologues had top hits with very negative scores, this suggests that they are highly likely to be OmpA-like proteins. Six of the top ten hits for BB\_0378 were RNA polymerase II proteins and only two of the top ten hits for BAPKO\_0387 were OM proteins. The high scores suggest that there is a greater than 3 % chance that the results are false positives, although the abundance of RNA polymerase II proteins in the top 10 results provides evidence that BAPKO\_0387, BB\_0378 and BG0377 are not related to the OmpA family. The top hit for both BB\_0543 and BG0553 was OmpX (one with a significant score, one with an insignificant score), however the vast majority of the rest of the hits were not OmpA-like proteins, and only one of the top ten hits for BAPKO\_0571 was an OmpA-like protein. This provides evidence for these proteins not being related to the family of OmpA-like proteins.

From FFAS03 analysis it appears that BAPKO\_0387, BB\_0378, BG0377, BAPKO\_0571, BB\_0543 and BG0533 were not OmpA-like TM -barrel proteins. Further analysis was carried out using signal peptide and TM -strand prediction.

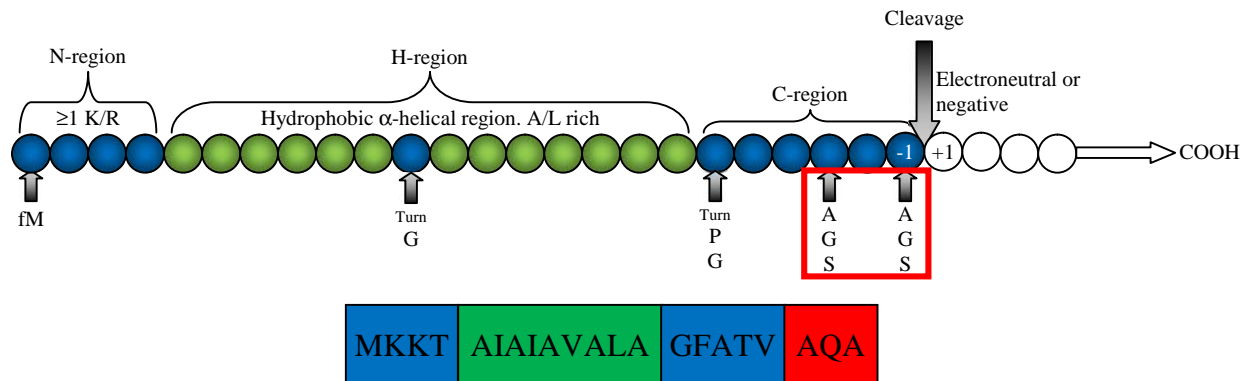
### 3.3.2 Signal peptide analysis

Signal peptides are present at the N-terminus of proteins that are functional at specific sites throughout the cell; the signal peptide facilitates translocation of the protein to its correct environment, e.g. the outer membrane of a bacteria (Pugsley, 1993) through various pathways. Each location around the cell has a different signal peptide, enabling recognition of the proteins function and required location (Pugsley, 1993). The standard signal peptide consists of three regions, N, H and C (Figure 3.7). The N region contains one or more positively charged residues, including one or more Lys or Arg and is between two and fifteen amino-acids in length. This is followed by a hydrophobic 'H' region with a high Aln/Leu content, but no charged residues (Arg, His, Lys, Asp or Glu) and no Pro. This region is generally greater than eight residues in length. The final region, designated as 'C' consists of residues that are less hydrophobic, and contains signals that are recognised by signal peptidases (LepA or LepB) (Figure 3.7).

Residues around the cleavage site are numbered according to position; with the cleavage site as the starting point and numbering outwards residues towards the N terminus of the protein are negative and positive towards the C terminus. Also visible is the 'A-A' box, present at the

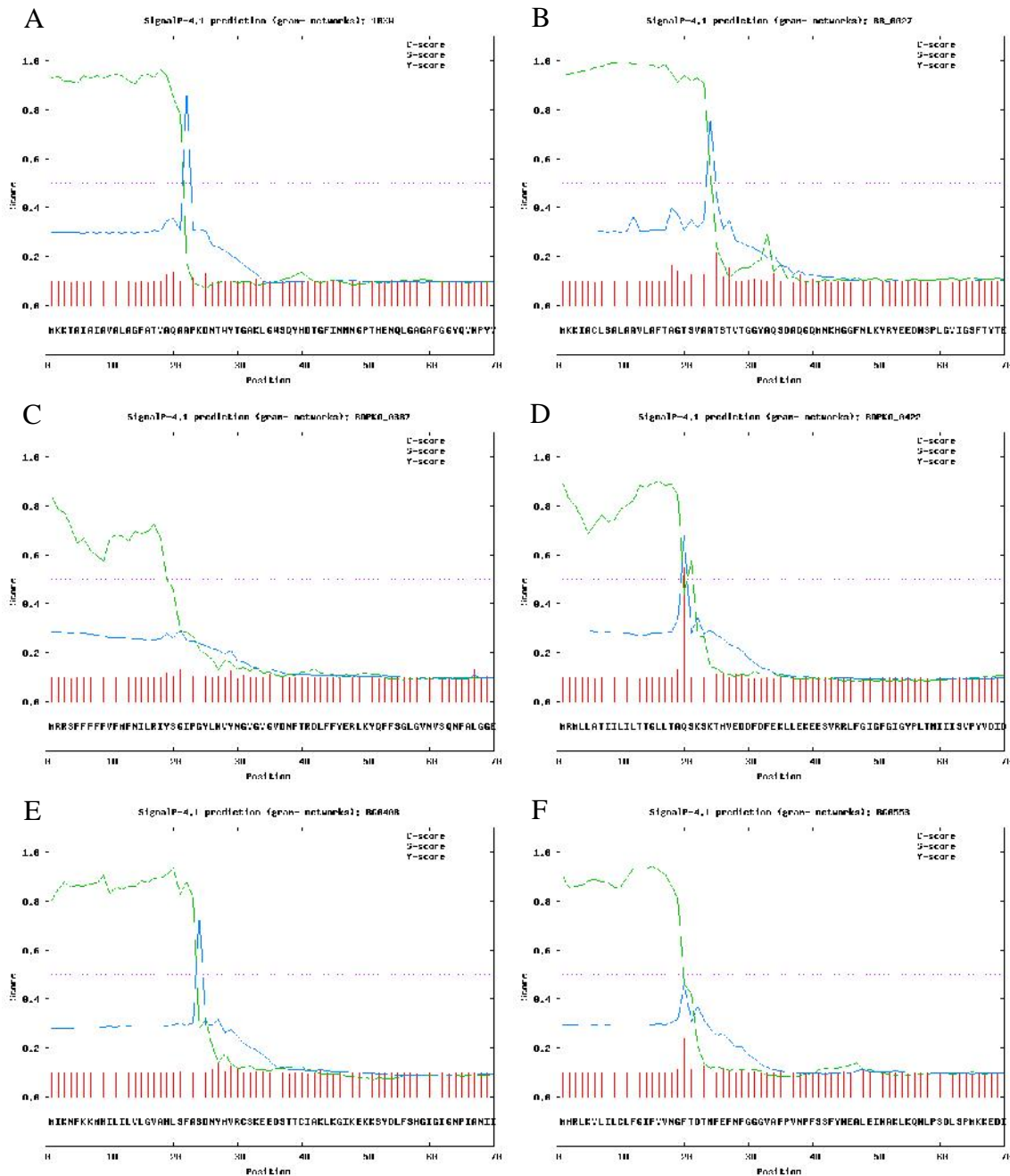


-1 to -3 positions on the amino-acid chain (Nielsen *et al.*, 1997). This is a common arrangement of residues in signal sequences in Gram-negative bacteria. The Aln residues may also be Gly or Ser and are considered important for the proper functioning of cleavage (Pugsley, 1993).



**Figure 3.7. Signal peptidases found in Gram-negative bacteria.** A typical arrangement of primary features of the signal peptide located at the N-terminus of a protein sequence. Dark blue coloured circles represent polar and charged residues, green circles represent hydrophobic residues. The N-region varies in size between 2 and 15 residues and has a net positive charge. The H-region is longer than 8 residues and generally does not contain any of the following residues P, K, R, D, E or H, but is usually alanine or leucine rich (Nielsen *et al.*, 1997). The C-region is shorter and less hydrophobic than the H-region. The A-A box (highlighted in red) is an important indicator of bacterial signal peptides (Nielsen *et al.*, 1997). Diagram adapted from (Pugsley, 1993). The boxed sequence displays the signal peptide from *E. coli* OmpA; L-R N-region (blue), H-region (green), C-region (blue) and the lipid box (red).

SignalP 4.1 (Petersen *et al.*, 2011) was used to predict the presence of signal peptides and the cleavage site of signal peptidases in putative *Borrelial* TM proteins (Table 3.3). *E. coli* OM proteins were used as positive controls. SignalP is designed to distinguish between signal sequences and α-helical transmembrane regions using two artificial neural networks, each ‘grown’ with the use of a different data set. Sequences containing transmembrane regions within the first seventy residues were used to train the SignalP-TM network, while the SignalP-noTM data set was trained without those sequences, instead using sequences of cytoplasmic and nuclear proteins (Petersen *et al.*, 2011). The network SignalP-noTM was used for prediction of signal peptides in the putative OmpA-like proteins. The option was present to select the type of bacteria (Gram-negative or positive) from which the sequences originate. As Spirochaetes are more closely related to Gram-negative bacteria than Gram-positive, this option was selected. The results of SignalP analysis along with positive controls (the N-terminal domain of *E. coli* OmpA in addition to OmpW and OmpX were submitted as positive controls) are displayed in Figure 3.8.



**Figure 3.8. SignalP 4.1 prediction of OmpA-like protein signal sequences.** A) 1BXW, B) BB\_0027, C) BAPKO\_0387, D) BAPKO\_0422, E) BG0408, F) BG0553. The graphical output of one of each protein homologues is included. Images generated by SignalP4.1 (Petersen *et al.*, 2011). The higher the c-score (red lines), the more greater the likelihood of a cleavage site at that position, the s-score (green line) represents the probability of the residue being part of the signal sequence. The y-score (blue line) is a combination of c-score and slope of s-score and is used to predict the cleavage site more accurately than the c-score alone (Petersen *et al.*, 2011).

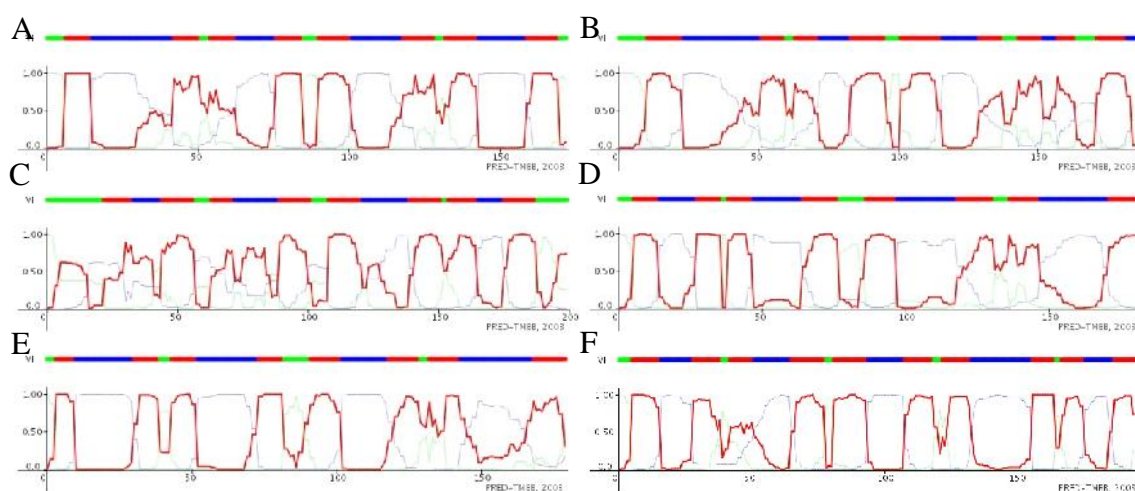
SignalP analysis found that BAPKO\_0387 and BB\_0378 are unlikely to contain signal peptides as the H-region is not A/L rich, and there is no ‘A-A’ box (Figure 3.8 C). This concurs with FFAS03 assessment that these are not TM -barrel proteins. In addition,



BAPKO\_0026 and BG0027 were also predicted to not contain signal peptides by SignalP (Table 3.6).

### 3.3.3 -barrel prediction

The amino-acid sequences of putative OmpA-like proteins identified by HMM searches were input into the webserver PRED-TMBB (Bagos *et al.*, 2004), which uses a HMM based method for predicting transmembrane -barrels in proteins (Figure 3.9). The N-terminal domain of *E. coli* OmpA in addition to OmpW and OmpX were submitted as positive controls.



**Figure 3.9. TM -barrel predictions of OmpA-like proteins.** A) 1BXW, B) BB\_0027, C) BAPKO\_0387, D) BAPKO\_0422, E) BG0408, F) BG0553. Data produced using the web-based server PRED-TMBB (Bagos *et al.*, 2004) using FASTA sequences with (the signal sequence as predicted by SignalP 4.1 has been removed) obtained from the RCSB PDB and BioCyc (Caspi *et al.*, 2010). Red signifies a predicted TM -strand, blue an extracellular loop and green a periplasmic turn. The score (from 0 to 1) relates to the probability of the region being membrane spanning. Most proteins were predicted to contain eight TM -strands, however BAPKO\_0571, BG0553 and BG0377 were predicted to have ten -strands and BG0027 to have nine.

The results of analysis by PRED-TMBB suggest that all proteins submitted are likely to be membrane-spanning -barrel proteins containing between eight and ten strands. BG0027 is predicted to have nine membrane-spanning -strands, which is unlikely as there are currently no known OM-spanning -barrel proteins with an odd number of strands. It is a possibility that BAPKO\_0571, BG0553 and BG0337 and BG0027 have ten TM -strands similar to the *E. coli* protein OmpT (Vandeputte-Rutten *et al.*, 2001).

### 3.3.4 Summary

Table 3.6 summarises the predictions from SignalP4.1 and PRED-TMBB

Protein	Position	SignalP 4.1		PRED-TMBB	
		D-score	Signal Peptide?	Score	-Barrel?
1BXW	1-21	0.906	YES	2.858	YES
2F1T	1-21	0.873	YES	2.902	YES
1QJ8	1-23	0.863	YES	2.816	YES
BAPKO_0026	1-24	0.504	NO	2.874	YES
BB_0027	1-23	0.863	YES	2.888	YES
BG0027	1-24	0.539	NO	2.876	YES
BAPKO_0387	1-20	0.467	NO	2.876	YES
BB_0378	1-20	0.544	NO	2.913	YES
BG0377	1-19	0.574	YES	2.908	YES
BAPKO_0422	1-19	0.742	YES	2.938	YES
BB_0405	1-18	0.727	YES	2.947	YES
BG0407	1-19	0.743	YES	2.938	YES
BAPKO_0423	1-23	0.752	YES	2.988	NO
BB_0406	1-23	0.819	YES	2.991	NO
BG0408	1-23	0.803	YES	2.989	NO
BAPKO_0571	1-19	0.681	YES	2.887	YES
BB_0543	1-21	0.821	YES	2.860	YES
BG0553	1-19	0.662	YES	2.912	YES

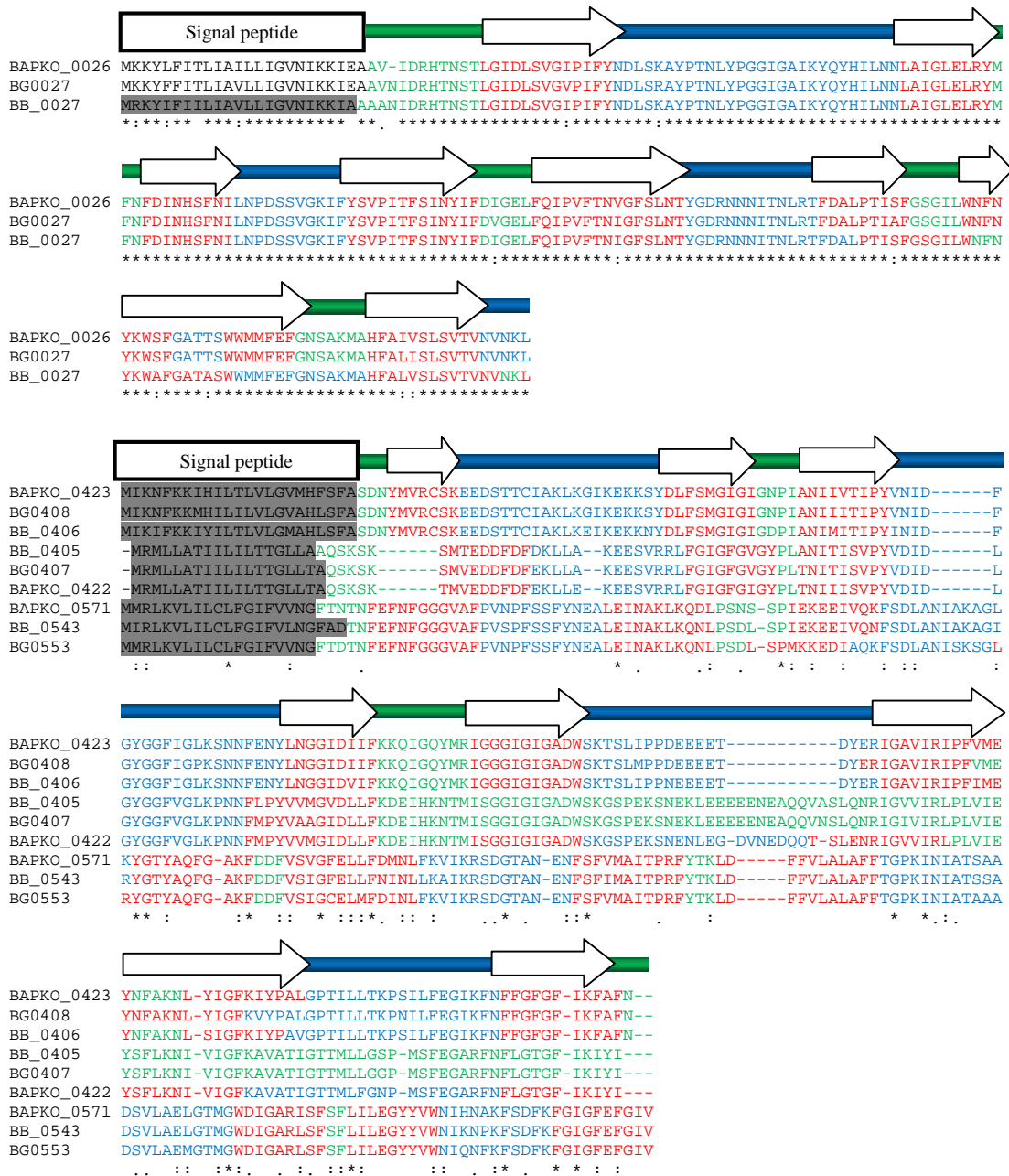
**Table 3.6. Results of SignalP 4.0 and PRED-TMBB analysis.** Three known proteins from *E. coli* are included for comparison. The position is the predicted position for the signal peptide. D-score is a weighted mean of the averaged S and maximum Y scores and is the distinguishing factor between signal and non-signal peptides, the cutoff for which was 0.570; scores above indicate a signal peptide (Petersen *et al.*, 2011). From the results obtained it was apparent that the majority of sequences contained a signal peptide. Scores are the discrimination score reported by PRED-TMBB. The threshold for the score is 2.965 and any proteins with a score lower than this are likely to be TM -barrels (reported as YES). All proteins with the exceptions of BB\_0406 and BG0408 were within this score and are therefore considered highly likely to be TM -barrel proteins.

BAPKO\_0387, BB\_0378 and BG0377 were discounted as potential OmpA-like proteins due to the FFAS03 prediction that they are homologous to RNA polymerase II, and the predicted absence of signal sequence indicative of cytoplasmic proteins (Figure 3.8 C). Putative OmpA-like proteins identified are identified below (Table 3.7).

<i>B. afzelii</i>	<i>B. burgdorferi</i> s.s	<i>B. garinii</i>
BAPKO_0026	BB_0027	BG0027
BAPKO_0422	BB_0405	BG0407
BAPKO_0423	BB_0406	BG0408
BAPKO_0571	BB_0543	BG0553

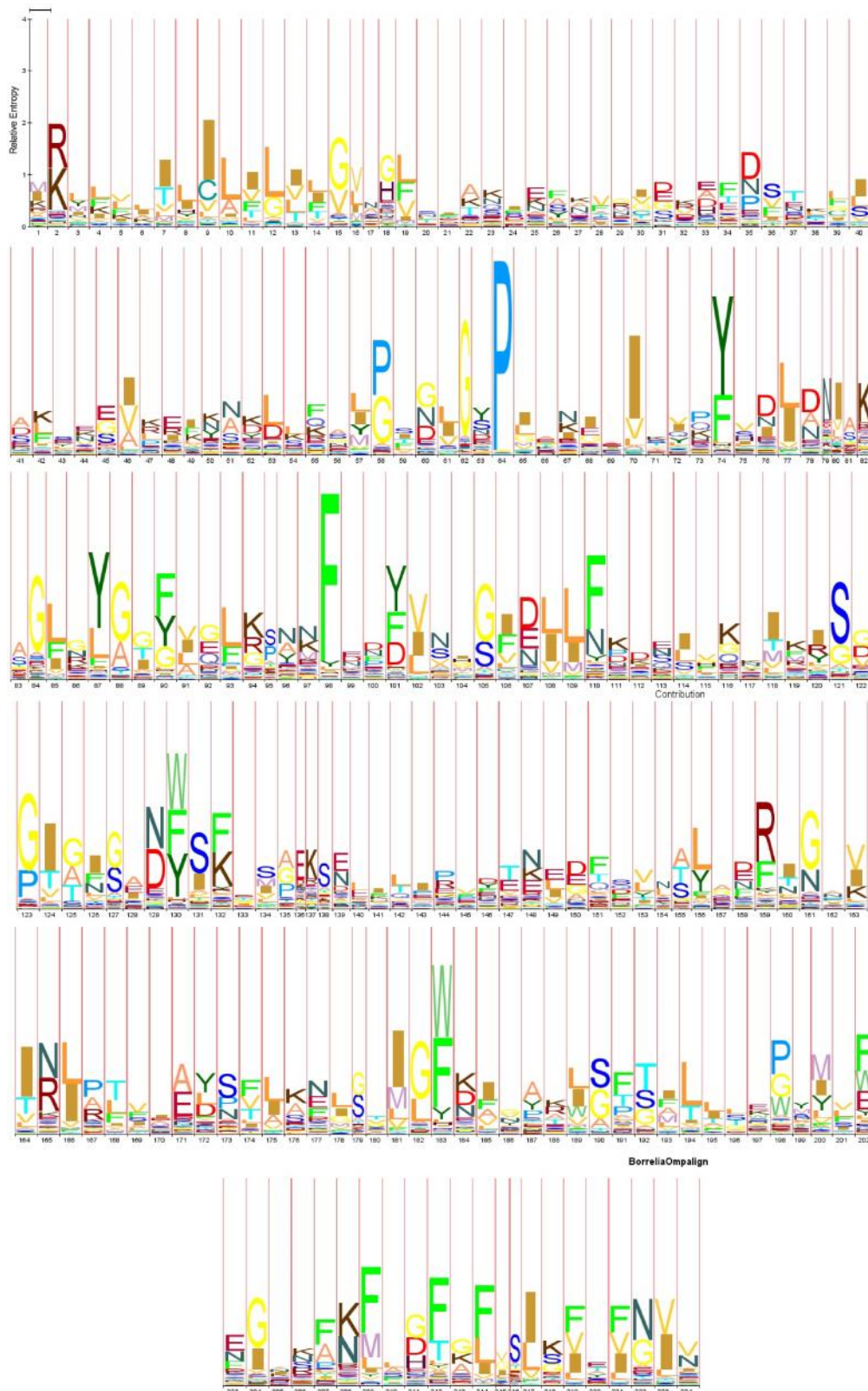
**Table 3.7. Putative small TM -barrel proteins in *Borrelia*.** These proteins were those identified as likely to be small TM -barrel proteins through the combined use of HMM profiling, FFAS03 assessment, SignalP analysis, and TMBB prediction.

A sequence alignment of putative *Borrelial* TM -barrel proteins (Table 3.7) was created (Figure 3.10) with the structural features predicted by SignalP 4.1 and PRED-TMBB highlighted. The signal peptide is highlighted gray, the periplasmic turns are green, TM -strands are red and extracellular loop regions are blue.



**Figure 3.10. Summary of topology prediction.** The sequences for four protein orthologues have been aligned using ClustalW2 (Larkin *et al.*, 2007). The signal sequence has been highlighted where there was predicted to be one (SignalP 4.1 (Petersen *et al.*, 2011) and the sequence coloured to match the PRED-TMBB predictions (Bagos *et al.*, 2004). BAPKO\_0026, BB\_0027 and BG0027 have been aligned separately as the sequence appears to differ more than the others. \* = a single fully conserved residue, : = conservation of strongly similar residues, . = conservation of weakly similar residues.

The sequence alignment of putative *Borrelial* OmpA-like proteins (Figure 3.10) was used to produce a HMM profile and HMM logo (Figure 3.11) for easy visualisation of residue conservation. There are a number of highly conserved aromatic residues which indicate the positions of the aromatic girdle. Towards the end of the HMM logo there are five tyrosine residues in quick succession, suggesting either mis-alignment of the sequences or a region rich in tyrosines. One proline residue is conserved almost entirely throughout the alignment, which is likely to be part of a periplasmic turn. As with the HMM logo of known OmpA-like proteins (Figure 3.6) there is a large number of conserved alternating glycine residues and long stretches of variable sequence, which again is indicative of the presence of TM  $\beta$ -strands and variable extracellular loops. Unlike the known TM  $\beta$ -barrel protein HMM logo (Figure 3.6) there is considerable conservation of several isoleucine and leucine residues, mainly towards the start of the sequence and a number of conserved phenylalanine residues near the C-terminus.



**Figure 3.11. HMM logo of putative *Borrelia* OmpA-like proteins.** Generated using LogoMat-M from a sequence alignment of identified putative *Borrelia* OmpA-like proteins (Figure 3.10). There are a number of highly conserved aromatic residues representing the aromatic girdle, in addition to stretches of alternating glycine residues characteristic of a T-strand and large variable stretches of residues indicative of extracellular loop regions.

### 3.4 Homology modelling of OmpA-like proteins

BAPKO\_0422 was modelled using the structures of the  $\beta$ -barrel domain of OmpA, OmpW and OmpX from *E. coli* as a template. The TM spanning regions from each of these proteins were identified along with the predicted TM regions of BAPKO\_0422 (PRED-TMBB) and aligned so that these regions matched as closely as possible. The loop regions of TM  $\beta$ -barrel proteins are highly variable both in length and sequence; thereby preventing meaningful sequence alignment. The TM spanning regions however are more consistent; being of comparable length, containing alternating hydrophobic and small side-chained residues and flanked by residues with bulky side-chains. These factors make for more readily identifiable regions and accurate alignments.

Using the sequence alignments (

Figure 3.12) models of BAPKO\_0422 were generated (Figure 3.13) using each of the three *E. coli* proteins as templates using Modeller v9.11 (Šali and Blundell, 1993).

```
>1BXW  -APKDNTWYTGAKLGWSQYHDTGLINNNGPTHENKLGAGAFGGYQVNPYVG--FEMGYDWLGRMPYKGSVENGAY
>0422  --QSKSKTMVEDDDFDEKLLK---EESVR----RLFGIGFGIGY-PLTNIIISVPYVDIDLGYGGFVGLKPNN

>1BXW  KAQGVQLTAKLGY-P-I-T-D-D-LDIYTRLGGMVWRAD-TYSNVYGKNHDTGVSPV-FAGGVEYA-I-T-P-EI
>0422  FMPYVVMGIDLLFKDEIHKNTMISGGIGIGADWSKGSPEKSNNLEGDVNEDQQTSLENRIGVVIRLPLVIEYSF

>1BXW  ATRLEYQWTNN-IG-DAH-TIGT-RP-DNG--MLSLGVSRYFG
>0422  LKNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI

>2F1T  -EAGEFFMRAGSATVRPTEGGGFSVTNNTQLGLTFTYMATDNIGVELLAATPFRHKIGTRATGDIATVHHLPT
>0422  QSKSKTMVEDDDFDEKLLK-EESVRRLFGIGFGIGY--P-LTNIIISVPYVDIDLGYGGFVGLKPNNFMPYVV

>2F1T  LMAQWYFGD-AS-SK-FRPYVGAGINYTTFFDNGFNDHGKEAGLSDLKDSWGAAGQVGVDY-LINR-DWLNV
>0422  MGIDLLFKDEIHKNTMISGGIGIGADWSKGSPEKSNNLEGD-VNEDQQ-TSLENRIGVVIRLPLVIEYSFLKN

>2F1T  MSVWYMDID-TTANYKLGAQQHDSVRLDPVVFMSAGYRFH
>0422  IVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI--

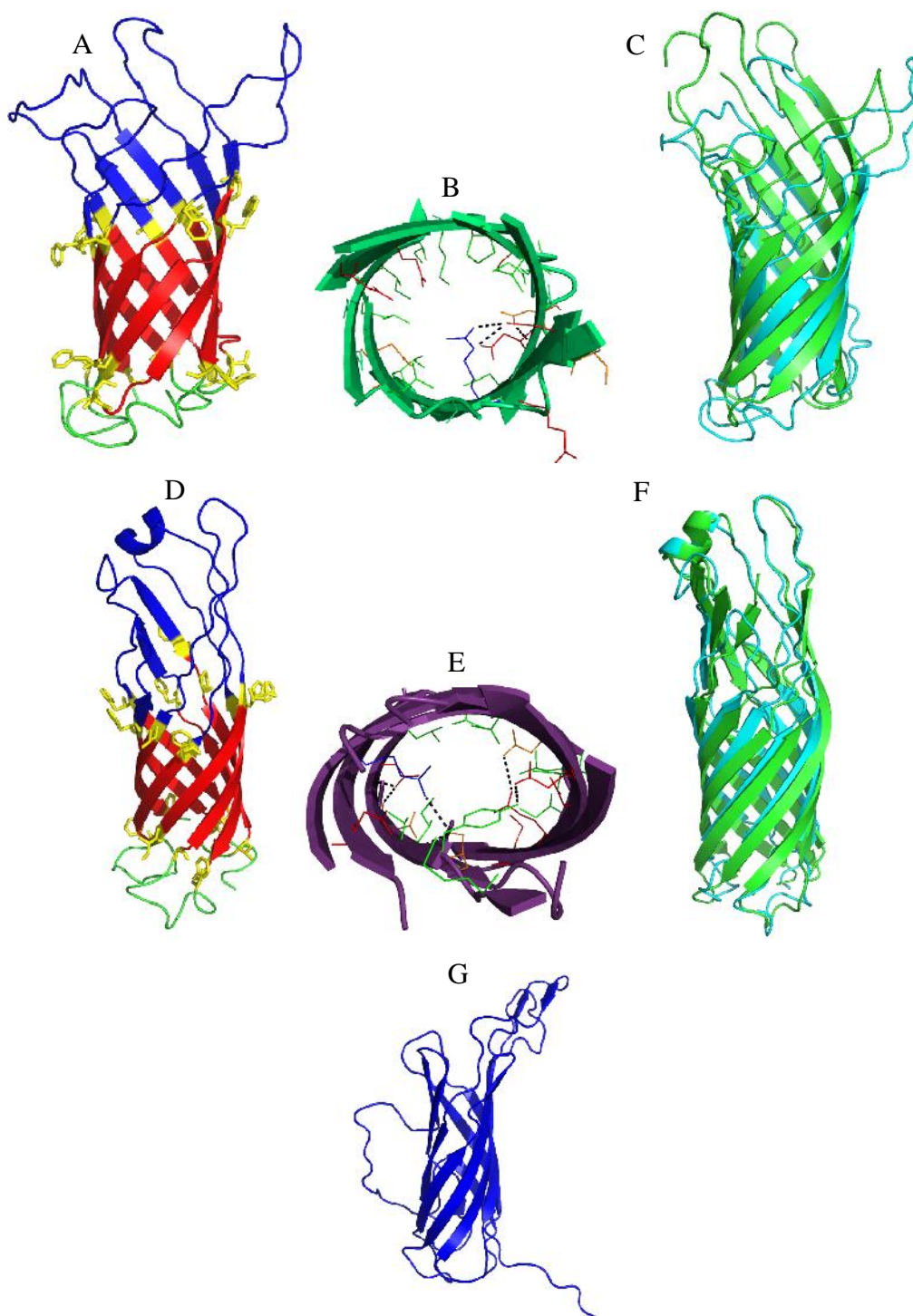
>1QJ8  -----ATSTVTGGYAQSDAQGQ-MNKMGGFNLKYRYEEDNSPLGVIGSFITYTEKSRTASSGDY--NKNQYYGITA
>0422  QSKSKTMVEDDDFDEKLLKEESVRRLFGIGFGIGY-PLTN-IIIISVPYVDIDLGYGGFVGLKPNNFMPYVVMGI

>1QJ8  GPAY-R-I-N-D-WA-SIYGVVGV-GYGK-FQTTEYPTYKNDT-SDYGFS---YGAGLQFNPMENVALDFSYSY---
>0422  DLLF--KDEIHKNTMISGGIGIGADWSKGSPEKSNNLEGDVNEDQQTSLENRIGVVIRLPLVIEYSFLKNIVIG

>1QJ8  ----E-Q-S-R-I-R-S-V-D--VGTWIAGVGYRF-*
>0422  FKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI*
```

**Figure 3.12. Sequence alignments of BAPKO\_0422 with known *E. coli* OM-spanning proteins.** Known TM spanning  $\beta$ -sheets are highlighted (green for OmpA, purple for OmpW and blue for OmpX) and predicted TM spanning  $\beta$ -sheets are highlighted red (BAPKO\_0422). These sequence alignments were used to generate homology models. Alignments in the .pir format used by modeller are included in appendix 8.2.





**Figure 3.13. BAPKO\_0422 Models produced using *E. coli* OM-spanning proteins as templates.** **A)** model of BAPKO\_0422 produced using 1BXW as template, **B)** view of internal H-bonding network of **A**, **C)** overlay of 1BXW (green) with **A)** (turquoise). **D)** model of BAPKO\_0422 produced using 2F1T as template, **E)** view of the internal H-bonding network in **D)**, **F)** overlay of 2F1T (green) with **D)** (turquoise). **G)** model of BAPKO\_0422 produced using 1QJ8 as a template. Models **A** and **D** have been coloured to show the periplasmic turns (green), extracellular loops (blue), membrane spanning regions (red) and aromatic girdle (yellow with side-chains shown).

The models produced using 1BWV and 2F1T as templates have longer periplasmic turns than expected compared to the known structures. Neither model shows the ‘gating’ residues present in OmpA (Smith *et al.*, 2007) and OmpW (Hong *et al.*, 2006) from *E. coli* (described in sections 1.5.1 and 1.5.2 respectively). The model based on 1QJ8 has a loop protruding from one of the strands and a long flexible region which would extend into the periplasm; in addition the loops are rather large, which provided evidence for BAPKO\_0422 not being an OmpX-type protein, or the sequence alignment could be in error. The down-the-barrel view of the H-bonding networks in models produced from 1BWV and 2F1T show minimal interactions. The H-bonding network in known OM-spanning proteins (Figure 1.29) is significantly more extensive than in the models. This may be due to the template having a different shear number to the *Borrelial* protein, meaning the barrels may actually be more compact than predicted in Figure 3.13.

### 3.5 Discussion

The overall sequence identity between *E. coli* OmpA/X and the *Borrelial* OmpA-like proteins identified in this study is very low (between 9-13%, Table 3.4), considered to be in the twilight zone (Rost, 1999). It is a general rule that for  $\beta$ -barrel formation, at least four of the five residues forming a  $\beta$ -sheet with the side chains oriented towards the membrane must be hydrophobic, with none of the three central residues charged, and that residues with side chains on the inside of the barrel must have small side chains (Koeblnik, 1999). The sequence alignment (Figure 3.4) of known OmpA (TM domains) and OmpX proteins shows high levels of conservation of alternating hydrophobic and small side-chained residues at regions thought to be membrane spanning; in particular glycines conserved with the side-chains protruding inwards from the barrel and aromatic residues (particularly tyrosine) at either end on the TM spanning domain that form the aromatic girdle. The conservation of residues is highlighted by the HMM-profile logo (Figure 3.6) where six tyrosines and several phenylalanines are visible as being highly conserved; these residues are likely to form the aromatic girdle surrounding the hydrophobic TM region. This conservation highlights the fact that the  $\beta$ -strands are important to the structure and orientation of the protein in the membrane. The sequence alignment and HMM-profile logo show that there are some exceptions to this rule as a glutamic acid residue appears to be present in the centre of one of the strands.



The final two  $\alpha$ -strands appear to be particularly well conserved, in particular the final three C-terminal residues: tyrosine, a positive residue (most likely arginine but potentially lysine) and phenylalanine, which are prominent in the HMM-profile logo (Figure 3.6). This combination of two aromatic and one positively charged residue forms the final three residues of a consensus sequence common to bacterial OM proteins identified by (Vogt and Schulz, 1999) - X-Z-X-Z-X-Z-X-Tyr-X-Phe (where X is any amino acid and Z is a nonpolar residue), which may be important for folding and insertion into the OM (Vogt and Schulz, 1999). Alternatively these residues may play a role in attachment and / or orientation of the N-terminal domain of OmpA. It is worth noting that the *Borrelial* proteins predicted to be OmpA-like do not have a conserved Y-K/R-F region at the C-terminus, suggesting that either the Y-K/R-F motif is not required in *Borrelia* or that these proteins are not  $\alpha$ -barrels.

In contrast to the conserved OM spanning regions, sequences of the loop regions of known OmpA-like proteins shows great variability in length and minimal conservation, in agreement with one of the 10 principles of  $\alpha$ -barrel building (Schulz, 2000); this indicates that the loop regions are not as important to the overall structure of the protein as the TM  $\alpha$ -strands are, and suggests that either they are not as important to the function of the protein, or that the proteins have slightly varying functions.

HMM searches are more sensitive than BLAST searches for distant homologues, but as a result produce more false positives, which requires further refinement of the results. For production of the HMM profiles, 23 and 68 sequences (from the 40 % and 90 % redundancy reductions respectively) were automatically aligned (ClustalW2). This automatic alignment does not take into account structural information such as membrane-spanning regions or turns/loops. Therefore key positions such as the aromatic girdle and alternating aliphatic residues may be misaligned; small errors in sequence alignment have a significant detrimental effect to the HMM profile, hence the very low scores obtained.

From the FFAS03 search results (Table 3.5) of putative *Borrelial* OmpA-like proteins several matches were obtained, each of which had a very low sequence identity (9 - 13 %) to known OmpA-like proteins. However this is comparable to the low sequence identity between *E. coli* and *Borrelial* BamA which are expected to have closely related structures and homologous functions. It is a possibility that both *Borrelia* and *E. coli* have these 8-stranded OM-spanning proteins because of their shared ancestry, or it could be that both bacterial

species have happened upon the same solution by convergent evolution. The conservation of a recognisable BamA export machinery may suggest conservation by shared ancestry. This would suggest that the 8-stranded OmpA-like proteins identified may have existed in an ancient common ancestor shared between *Borrelia* and Gram-negative bacteria. The low sequence identity also provides an explanation for the lack of hits obtained by BLAST searching using *E. coli* OmpA as a target.

Five proteins identified by HMM searches had a top FFAS hit with the score less significant than the cutoff value, meaning that there was a higher than 3 % chance of false results (Table 3.5); for BAPKO\_0387, BB\_0378 and BG0377 the FFAS scores for the top ten hits were all insignificant, and six of these were an RNA polymerase II protein; meaning it highly likely that these proteins were not OmpA-like. The two homologues to BAPKO\_0422 both scored lower than the cutoff and all three had top hits of OmpA-like proteins, BAPKO\_0571 and BB\_0543 had top hits of OmpX (though less significant than the score of -9.75) and a mixture of hits for the rest of the top ten. BAPKO\_0026, BB\_0027 and BG0027 each had scores 3-4 times more significant than the cutoff value, representing a high probability that the OmpA-like protein matches found were accurate. These proteins appeared highly likely to be OmpA-like proteins in addition to BAPKO\_0423, BB\_0405, BB\_0406, BG0407 and BG0408 (all with scores more significant than the cutoff).

Analysis of the putative OmpA-like proteins for a signal sequence revealed that four proteins were predicted to not have signal peptides, compared to eleven that were. It is worth noting that the network trained to determine signal peptides and cleavage point is designed for Gram-negative bacteria. Spirochaetes have a double membrane morphology reminiscent of Gram-negative bacteria, however there are several important differences in the structure and chemical composition of the *Borrelial* membrane, for example *Borrelia* do not express LPS and are known to contain lipid rafts. Although the SecA-dependant secretory pathway is highly conserved among Gram-negative bacteria and Spirochaetes, it remains a possibility that some sequence divergence may have occurred in *Borrelia* reducing the efficiency of prediction algorithms such as SignalP, which explains why not all signal sequences are predicted (Table 3.6). Supporting this, the Spirochaetal lipoprotein signal sequence was shown to be significantly different to other bacteria (Setubal *et al.*, 2006). The Gram-negative algorithm implemented by SignalP is based on a training set of 266 experimentally confirmed

secreted proteins (Nielsen *et al.*, 1997, Petersen *et al.*, 2011). SignalP has a reliability of 0.948 for proteins with no TM regions (  $\alpha$ -helices) (Petersen *et al.*, 2011).

The four proteins predicted to not contain a signal sequence were BAPKO\_0026, BG0027, BAPKO\_0387 and BB\_0378. It was interesting that BAPKO\_0026 and BG0027 were predicted to have no signal peptide as these proteins were predicted the most likely to be OmpA-like from FFAS03 searching. The graphical output for these proteins (not shown) showed that the S-score was representative of a signal peptide; however a cleavage site (C-score) could not accurately be determined, resulting in a low Y-score and the resulting prediction of no signal peptide. The amino-acid sequence shows one common polymorphism in BAPKO\_0026 and BG0027 from that of BB\_0027; at position 23 in BB\_0027 is an alanine residue, which in conjunction with the two following alanines forms an 'A-A' box (Figure 3.7) (Pugsley, 1993). Residue 23 in BAPKO\_0026 and BG0027 is a glutamic Acid, whilst residue 26 is a valine, resulting in no 'A-A' box. The sequence is EAAV; a mutation of either the glutamic acid or the valine to an alanine would result in an 'A-A' box and the recognition of a signal peptide. It is possible that the *Borrelial* machinery recognises the non-standard 'A-A' box, or that this could be a pseudo gene in *B. afzelii* and *B. garinii*. *Borrelia* contains a SecA homologue responsible for OM protein export that has only 35.0 % to 43.2 % sequence identity to other bacterial homologues, hence it is likely that at least some components and mechanisms of protein secretion and its regulation will be novel (Guina *et al.*, 1998). Although there are a large number of pseudo genes on the plasmids, the linear chromosome is not known to contain a significant proportion of pseudo genes. Experimental expression of BB\_0027 (numbered BB\_0028 in the publication due to being expressed from a different isolate) and BB\_0405 has been shown in OM vesicles of *B. burgdorferi* (Yang *et al.*, 2011), providing evidence that this sequence is a functional signal peptide recognised by *Borrelial* machinery and the protein is exported to the OM.

Prediction of transmembrane  $\alpha$ -barrel proteins predicted that all proteins with the exception of BAPKO\_0023, BB\_0406 and BG0408 were transmembrane  $\alpha$ -barrel proteins. The score generated by these proteins was above the cutoff for classification as TM  $\alpha$ -barrel proteins, though this is in contrast to the graphical output for these three proteins, which clearly showed eight membrane spanning domains, four extracellular loops and three periplasmic turns (Figure 3.9 D). PRED-TMBB, like SignalP 4.1 is trained on Gram-negative proteins and so it is possible that this was a false negative result. Visible in the sequence alignment

(Figure 3.10) are aromatic residues present at least one end of each predicted TM spanning  $\beta$ -sheet, these proteins would form the aromatic girdle, which is characteristic of TM  $\beta$ -barrel proteins (Schulz, 2000). Some of the proteins predicted to be transmembrane  $\beta$ -barrel proteins were predicted to have 9 strands (BAPKO\_0026 and BG0027) (Figure 3.9 B & Figure 3.10) which according to the principles of  $\beta$ -barrel proteins set out (Schulz, 2000) is highly unlikely, no known OM proteins have an odd number of strands. An odd number of  $\beta$ -strands would be unlikely from a biological standpoint as an even number of strands is likely to be a restraint of the folding mechanism. Observation of the amino-acid sequences and comparison with the BB\_0027 which was predicted to have eight strands shows that the change to amino-acid sequence is exactly the same in this area in all three sequences, with three minor exceptions: positions 162, 176 and 181. In BAPKO\_0026 and BG0027, position 162 is a serine and alanine respectively and a serine in BB\_0027; this appears to have had no affect on the prediction as two different residues produce the same strand. The other two polymorphisms are serine (nine strands) to alanine (eight strands) and threonine (nine strands) to alanine (eight strands). It is possible that the prediction algorithm is very sensitive to changes in sequence. Upon looking at the sequence (Figure 3.9) it would appear that the strands predicted for both eight and nine stranded proteins are viable, however the eight stranded prediction is more likely as the strand size is consistent, as are the alternative hydrophobic side-chains. Interestingly, the previous version of PRED-TMBB predicted eight strands in all three of these proteins (data not shown). The graphical output (Figure 3.9 B) shows a slight rise in the probability of a  $\beta$ -strand between the first two predicted strands. Looking at the sequence (Figure 3.10) shows a possible TM  $\beta$ -sheet region flanked by aromatic residues, however if this was a  $\beta$ -strand, all predicted turns would become loops and vice versa, resulting in very short loops and long turns. According to the ten principles of  $\beta$ -barrel building (Schulz, 2000) turns should only be a couple of residues in length, and loops much longer, hence it is much more likely that the protein is eight stranded and the program has predicted a turn in the middle of a strand, resulting in two short strands where only one should be.

Proteins BAPKO\_0571, BB\_0543 and BG0553 were predicted to contain ten  $\beta$ -strands each. This is possible as this is an even number of strands which matches the principles of  $\beta$ -barrel construction (Schulz, 2000) and many  $\beta$ -barrel proteins are known to contain more than eight strands. *E. coli* is known to contain a 300 residue ten stranded TM  $\beta$ -barrel known as OmpT (Vandeputte-Rutten *et al.*, 2001) and a twelve stranded TM  $\beta$ -barrel protein - Tsx, a

nucleotide transporter protein (Hong *et al.*, 2006). The length of the loop regions appears shorter than other putative OmpA-like proteins; analysis of the sequence (Figure 3.10) shows a number of aromatic residues where the aromatic girdle would be expected and residues with small side-chains present on the inside of the barrel, however the protein sequence is approximately 100 amino-acids shorter than the sequence of OmpT, making it unlikely that the protein is ten stranded.

Models produced using Modeller v9.11 (Figure 3.13) are consistent with BAPKO\_0422 being a transmembrane  $\beta$ -barrel protein, most likely an OmpA-like or OmpW-like protein. There is a clear aromatic girdle present around either end of the predicted TM-spanning  $\beta$ -barrel in addition to a large number of glycine residues present on the interior of the barrel. The girdle in the predicted models are not completely composed of aromatic residues; nine of sixteen residues are either tryptophan, phenylalanine or tyrosine. This is in approximate agreement with OmpW from *E. coli*, which has eleven aromatic residues, and matches the nine aromatic residues in OmpA from *E. coli* exactly, thus providing good evidence for the accuracy of the model and the presence of an aromatic girdle within BAPKO\_0422. The models do, however show a discrepancy with the length of periplasmic turns; sometimes up to several residues longer in the *Borrelial* protein.

BAPKO\_0422 appears to have a peptide sequence too long to be similar to that of OmpX (182 residues compared to 148), both the periplasmic turns and extracellular loops are much longer than *E. coli* OmpX and when modelling unordered loops have a tendency to form in the middle of a predicted  $\beta$ -strand. From the models produced it is apparent that the putative OmpA-like proteins in *Borrelia* have TM spanning regions of similar length to those present in *E. coli*. The homology models based on OmpA and OmpW have significantly fewer internal H-bonds than OmpA or OmpW. This could be a result of a number of factors such as the sequence in the models not being aligned correctly, the strands being too far apart for the software to register interactions or incorrect prediction of the sheer number. It is also possible that these proteins, while similar in structure to *E. coli* OmpA have different functions, such as performing as pores. The sequences are different to those of *E. coli* OmpA-like proteins and as such it is not guaranteed that the internal structures of the proteins would be similar.

Bioinformatic analysis suggests that the OmpA-like proteins identified in *Borrelia* are more likely to be OmpX or OmpW than OmpA. OmpA in Gram negative bacteria is known to

interact with the peptidoglycan layer present just beneath the OM, however the putative - barrels identified in *Borrelia* lack such a domain and the peptidoglycan layer is located much closer to the cytoplasmic membrane. The inside of the barrel of BAPKO\_0422 is predicted to contain many glycine residues and resemble an inverse micelle, similar to OmpX, however the loop regions are predicted to be much longer than OmpX, and more characteristic of OmpW. As OmpW is responsible for the passive uptake of hydrophobic nutrients (Hong *et al.*, 2006), not all bacteria have a requirement for this protein, in which case *Borrelia* may contain OmpW or a related protein.

## Chapter 4. Cloning and preparation of OmpA-like protein for structural studies

### 4.1 Amplification of genes for cloning

Genes encoding for putative OmpA-like proteins in *B. afzelii*; BAPKO\_0026 and BAPKO\_0422, and *B. garinii*; BG0027 and BG0407 were amplified using nested PCR and the primers in Table 2.5 & Table 2.6. No *B. burgdorferi* s.s. genes were amplified as no *B. burgdorferi* s.s. DNA had been sourced at the time of amplification.

In a standard PCR reaction, there are three stages; initial denaturation, amplification cycle and final extension. The initial denaturation step is the heating of the reaction mix to 95 °C for five minutes. This denatures (or ‘unzips’) double stranded DNA, exposing the base groups on each strand. Immediately after this amplification cycles commence. Each amplification cycle consists of three stages; denaturation, annealing and extension.

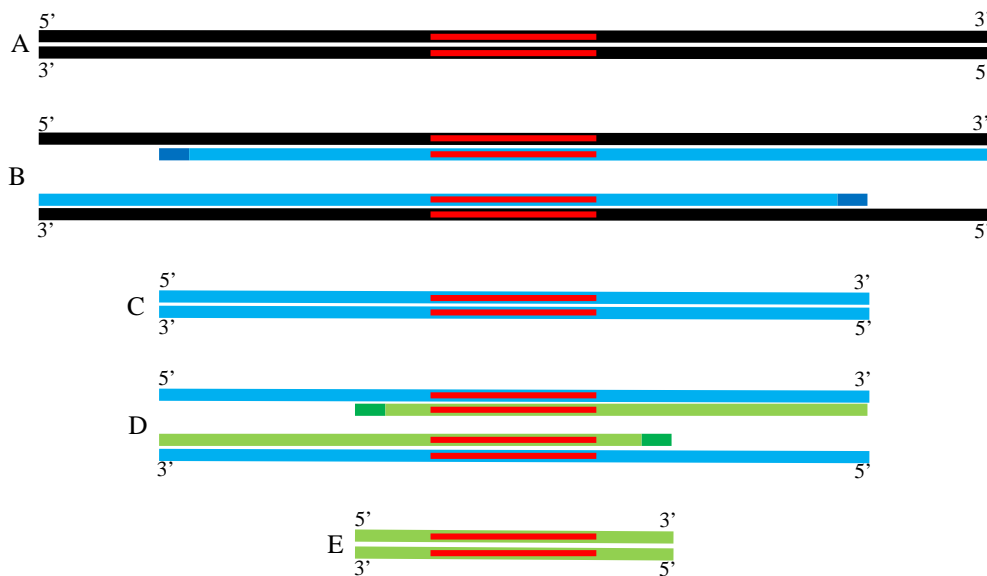
- Denaturation – same as initial denaturation, but for thirty seconds rather than five minutes
- Annealing – the reaction mix is cooled to a temperature a couple of degrees below the lowest melting temperature ( $T_m$ ) of the primers to enable primer binding
- Extension – the reaction mix is heated to 72 °C to enable Taq DNA polymerase extension of the primers using the initial strand as a template

Standard PCR procedure is to repeat this amplification cycle twenty eight times before the final extension stage of five minutes at 72 °C.

Nested PCR is a highly sensitive technique that uses two consecutive stages of PCR to amplify the target region of DNA with a limit of detection comparable to that of real-time PCR (Chacón and Ferreira, 2008). This enables amplification of a specific gene from a low concentration sample in the presence of a high DNA contamination, in this case *Borrelial* DNA in the presence of *Ixodid* tick DNA. The initial stage of nested PCR amplifies a fragment of DNA slightly larger than the required target gene, the second stage then amplifies the target gene from within the initial amplification (Chacón and Ferreira, 2008). The amplification of two or more regions as carried out in nested PCR has been shown to produce a higher sensitivity than when amplifying one region as in traditional PCR (Grankvist *et al.*, 1996, Zhang *et al.*, 2007). The use of two primer pairs greatly reduces the likelihood of amplification of an incorrect gene; the first round of amplification is susceptible

to low-level amplification of non-specific gene fragments, but it is highly unlikely that these fragments would be amplified by the second primer pair. The amplification process is summarised in Figure 4.1

As nested PCR (nPCR) is at its simplest two full rounds of PCR using different primers, the major difference is the number of cycles in the amplification phase. The annealing temperature is varied depending on the  $T_m$  of the primers. The amplification stage is repeated twenty times in the initial PCR using initial (or ‘outer’) primers and thirty five times in the secondary PCR using secondary (or ‘inner’) primers.

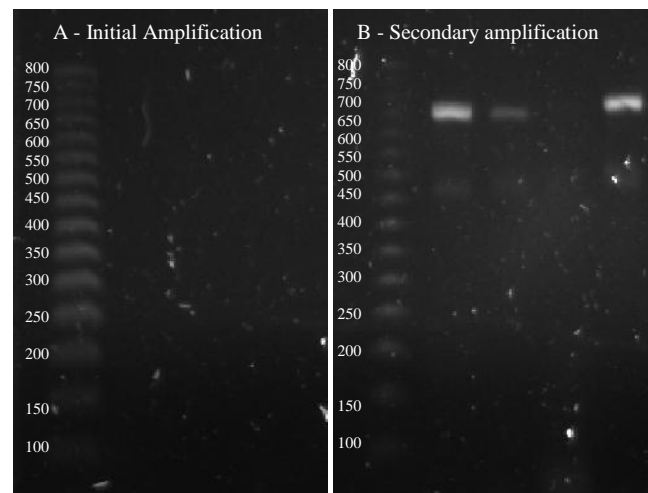


**Figure 4.1. The amplification of a target gene by nested PCR.** **A)** double stranded genomic DNA (gDNA) with target gene highlighted in red. **B)** Initial amplification: gDNA unzipped with primers bound (dark blue) and extended (light blue) using gDNA as a template. **C)** DNA fragment comprising target gene and a large surrounding area. **D)** Secondary amplification: Initial fragment has been unzipped and secondary primers (dark green) bound and extended (light green). **E)** Final gene fragments. Once a strand has been replicated, this strand can then be used as the template for replication in the opposite direction. The replication is cut short when the end of the strand is reached, thus producing a fragment. The result of the amplification is a large number of fragments, the majority of which are concise fragments. The process is repeated again, this time amplifying the target gene and only a small surrounding area, mainly from the synthesised fragments (light blue). This step produces a large quantity of target gene (light green and red).

Housekeeping genes are required for maintenance of basic cellular function, and as such are employed as positive controls for PCR amplifications. Housekeeping genes PepX and UvrA (section 2.1.2) were selected and amplified using the primers in Table 2.4. BG0027 was selected as an experimental amplification. For BG0027 reactions 1 & 2 the outer primer pairing was G0027OF1/G0027OF2, and the secondary (inner) primer pairings were G0027IF1/G0027IR1 and G0027IF2/G0027IR2 respectively (Table 2.5). The primary

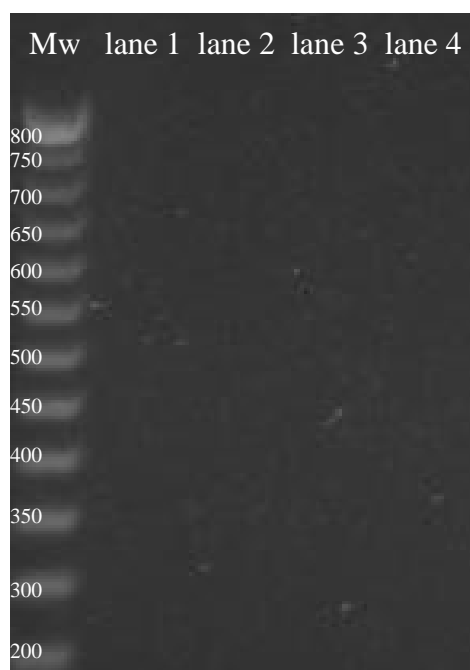


amplification used the annealing temperature 50 °C (housekeeping genes) and 60 °C (BG0027), whilst all secondary amplification were carried out with an annealing temperature of 52 °C. There was visible product of approximately the correct size (~660 bp) in both housekeeping genes along with one of the target gene combinations (Figure 4.2), demonstrating that the nested PCR method was effective. All gel analysis of PCR products was carried out using a 2 % agarose gel as described in section 2.2.8.2.



**Figure 4.2. Amplification of PepX, UvrA and BG0027 using nPCR.** A) Lanes from L-R: 800 bp step ladder, PepX, UvrA, BG0027 (1), BG0027 (2). B) L-R: 800 bp step ladder, PepX, UvrA, BG0027 (1), BG0027 (2)., BG0027 was selected at random for experimental method testing. Three bright product bands are visible at approximately 650 – 700 bp. In addition a faint, smaller PCR product at around 450 bp is visible in all three reactions. These are likely bands produced by non-specific primer binding.

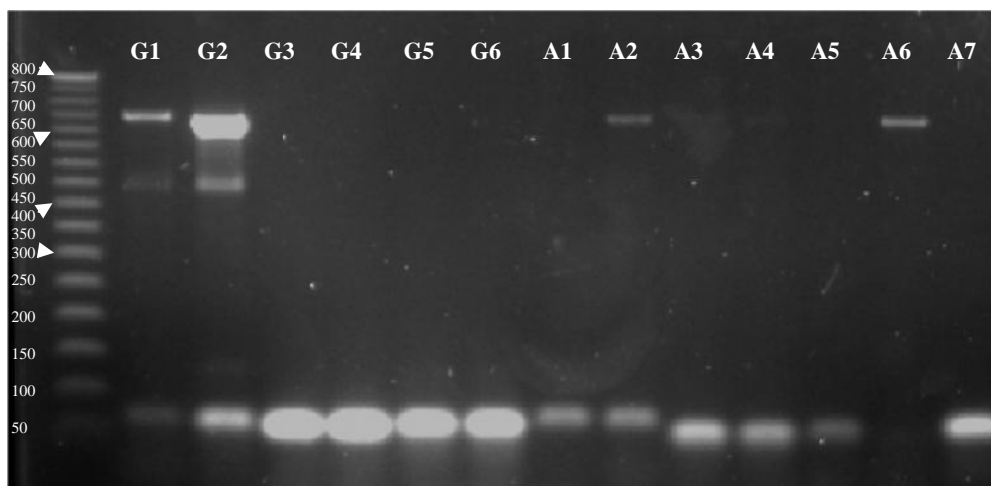
The absence of product in the initial amplification provided evidence that nPCR is required to produce sufficient product. The failed amplification from reaction 1 for BG0027 indicates that there was likely a problem with the primer  $T_m$  difference. Single-step PCR was attempted using the secondary amplification pairs was conducted on the housekeeping genes and BG0027 for the same number of inner amplification cycles (35) as nPCR, however there were no products produced (Figure 4.3) confirming that nested PCR was required.



**Figure 4.3. Gel of 35 cycle PCR using secondary primers only.** Mw is the 800 bp step ladder, lanes 1-4 are as follows: PepX, UvrA, BG0027 (1), BG0027 (2). PCR conducted using an annealing temperature of 50 °C (housekeeping genes) and 60 °C (BG0027). Primers used were those used for secondary amplification previously (Figure 4.2 B) No PCR product was produced from either of the housekeeping genes or the experimental amplifications.

Amplification of the orthologous gene BAPKO\_0026 was carried out in the same manner, in this case, primer pairs A0026OF1/A0026OR1 and A0026IF1/A0026IR1 yielded sufficient amplification of the gene; though the annealing temperature had to be adjusted to 56 °C (initial) and 59 °C (secondary) for optimum amplification.

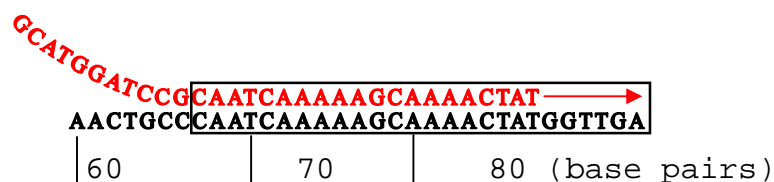
BAPKO\_0422 and BG0407 were not successfully amplified using initial primer combinations and annealing temperatures, hence a broad screen was conducted using multiple combinations of primary and secondary primers (Table 2.5) for each gene (Figure 4.4). Primer pairs with a  $T_m$  difference greater than 5 °C were not used. In all, six different primer combinations were used for BG0407 and seven for BAPKO\_0422. The annealing step temperatures selected for this trial were 51 °C for the initial amplification and 59 °C for the secondary amplification.



**Figure 4.4. Amplification screen for BG0407 and BAPKO\_0422.** Far left; the standard 50 bp step ladder (NEB) used. Lanes G1-G6 are the amplification attempts of BG0407, lanes A1-A7 are BAPKO\_0422. Primer combinations G1 and G2 from BG0027 and A2, A4 and A6 from BAPKO\_0422 have produced PCR products. Both BG0407 reactions have produced two PCR products, one of which is approximately 500 bp. For the addition of restriction sites, PCR product from primer combinations G1 and A6 were chosen. In addition, G2 appeared to have produced two bands of very similar size at approximately 600/650 bp. The larger band produced from G1 was excised, purified and this used for production of constructs.

Successful primer combinations for the amplification of BG0407 were: G0407OF1/ G0407OR1 with G0407IF1/ G0407IR2 (G1) and G0407OF1/ G0407OR1 with G0407IF2/ G0407IR1 (G2) and for BAPKO\_0422: A0422OF1/ A0422OR1 with A0422IF2/ A0422IR1 (A2); A0422OF1/ A0422OR2 with A0422IF2/ A0422IR1 (A4) and A0422OF2/ A0422OR2 with A0422IF2/ A0422IR1 (A6). Combinations G1 (the product at approximately 660 bp, excised and purified) and A6 were selected as G1 produced the clearest amplification of BG0407, whereas G2 appeared to have produced two fragments of very similar size as evidenced by the bright bands at approximately 600 bp. Both A2 and A6 produced clear individual bands, however as A6 was brightest, this was chosen. There is a very faint and present from primer combination A4.

The product was purified and restriction sites were added to each end of the gene through a further 28 cycle PCR using overhang primers (Table 2.6) as detailed in section 2.2.1. Constructs were designed to produce a recombinant protein resembling the native protein as close as possible. In order to achieve this, primers were designed to start the recombinant protein at residues immediately following the predicted signal peptide cleavage site. Primers for adding restriction sites were designed with a small region of DNA that was designed to ‘overhang’ upon annealing to the target strand (Figure 4.5).



**Figure 4.5. An example the positioning of a forward overhang primer, A0422CF.** In black text is the PCR product after amplification of the gene with both initial and secondary primers. In red is the overhang primer, the black box is the amplified product and the red arrow is the direction of amplification. The numbering of base pairs begins at the beginning of the gene. Positioning of the primer in this way removes the signal peptide.

The majority of the primer matches with the DNA strand (and so will bind to the complementary strand) however, the initial 11 bases do not. After amplification beginning from this primer is complete, the reverse primer will bind to the amplified strand and amplification will proceed along using the newly synthesised strand as a template, including the former overhang. This will produce a complete restriction site ready for further amplification and digestion. The primer was designed to overlap the start of the gene, minimising the number of extra residues at the start of the protein, in some cases, extra bases had to be added between the restriction site and the start codon in order to bring the sequence into frame. PCR products with restriction sites were sent for sequencing (Source Bioscience, Nottingham, UK). An example of sequencing results for BAPKO\_0422 follows (Figure 4.6)

```
>BAPKO_0422_F
NNNNNNNANNNAACANCAGGCTGTGTTACTGCCCAATCAAAAAGCAAATCTATGGTTGAAGATGACTTTGACTTTGAGAACTTCTTGCAAAA
GAAGAGTCTGTCCGCCGTTTATTTGGTATAGGCTTTGGGATTGGATACCCACTTACAAACATTACAATATCTGTTCCATATGTAGACATAG
ACCTTGGCTACGGAGGATTTCGTTGGGCTTAAACCCAACAATTTTCATGCCCTATGTTGTAATGGGAATAGACCTTTTATTTAAAGATGAAAT
ACATAAAAACACCATGATTTCGGCGGCATTGGAATAGGAGCAGATTGGTCAAAGGAAGCCCTGAAAAATCGAATGAAAACCTTGAAGAA
GACGAACAAGAAGATCCTCAACATCTCTTTCTCATCAAAATAGAAATAGGGGTGTGATACCTTTGNCTTTGGAATCAAATTTCTCCCTA
ATAAAAAAATTGGGGTTAAAGCTGTTGCTNCTANTGNAACTACNACAACATTTGNNNNNNCNCNTGCCATTTGNATTNGCNGNANNTANTTT
CNNGTNCACGGGCNTNNGNNNTATATATATATATNGNNNNNNNANNNNNANNTCCNNNAANNANAAAN
```

```
XXXNXRLVTAQSKSKSMVEDDFEFKLLAKEESVRRLFGLGFGIGYPLTNITISVPYVDIDLGYGGFVGLKPNFMPYVVMGIDLLFKDEI
HKNTMISGGIGIGADWSKGSPEKSNENLEEDEQEDPQTSLSHQNRIGVVIPLXLESNFLPNKKIGVKAVAXXTXTTFXXXXPFXXXXXXF
XXTGXXXIYIYXXXXXXSXXXXX
```

**Figure 4.6. Results of BAPKO\_0422 sequencing.** Top - DNA was sequenced in the 5'-3' direction on both the forward and reverse strands using BA0422CF as primer. Some bases were not able to be automatically assigned and so were assigned manually where possible. Bases that were unassigned are denoted as N. Bottom - Translation of sequenced DNA, the first three residues of the protein sequence are highlighted grey and the last three light grey. Residues unable to be assigned are denoted as X.

The gene sequence was translated to the amino-acid sequence using the ExPASy Translate tool (Figure 4.7). This was BLAST searched to ensure that the correct gene had been amplified.

tr	Q0SNA2 Q0SNA2_BORAP	Full=Uncharacterized protein; [Ordered Locus Names = BafPKo_0408, BAPKO_0422;] [Borrelia afzelii (strain PKo)]	201 AA
Score = 354 bits (909), Expect = 5e-96 Identities = 177/182 (97%), Positives = 178/182 (97%)			
Query: 1	QSKSKTMVEDDDFEKLLLEKEESVRRLLFGIGFGIGYPLTNIISVAAYVDIDLGYGRFVGL	60	
	QSKSKTMVEDDDFEKLLLEKEESVRRLLFGIGFGIGYPLTNI ISV YVDIDLGYG FVGL		
Sbjct: 20	QSKSKTMVEDDDFEKLLLEKEESVRRLLFGIGFGIGYPLTNIISVPIYVDIDLGYGGRFVGL	79	
Query: 61	KPNNFMPYVVMGIDLLIKDEIHKNTMISGGIGMGADWSKGSPEKSSENLEGDVNEDQQTS	120	
	KPNNFMPYVVMGIDLL KDEIHKNTMISGGIG+GADWSKGSPEKSSENLEGDVNEDQQTS		
Sbjct: 80	KPNNFMPYVVMGIDLLIKDEIHKNTMISGGIGMGADWSKGSPEKSSENLEGDVNEDQQTS	139	
Query: 121	LENRIGVVIRLPLVIEYSFLKNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKI	180	
	LENRIGVVIRLPLVIEYSFLKNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKI		
Sbjct: 140	LENRIGVVIRLPLVIEYSFLKNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKI	199	
Query: 181	YI	182	
	YI		
Sbjct: 200	YI	201	

**Figure 4.7. Amplified BAPKO\_0422 BLAST result.** Displayed is the BLAST result of translated gene searched against the *B. afzelii* proteome showing the top hit Q0SNA2, which from the bioinformatic results (chapter 3) is the putative protein. Q0SNA2 (also noted as BafPKo\_0408 and BAPKO) There are five polymorphisms (turquoise) in the amplified gene compared to that in the database.

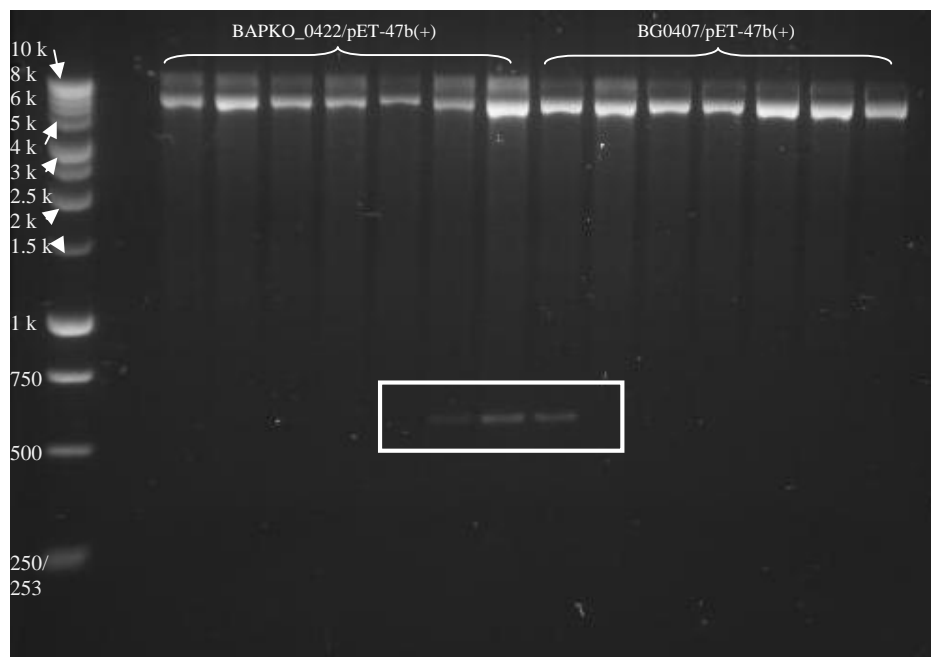
Five polymorphisms were visible in the gene compared to the peptide sequence in the UniProtKB database. The database sequence was sequenced from laboratory cultured bacteria, whereas the gene amplified was from wild-type *Borrelia*.

## 4.2 Production of gene constructs

The amplified genes and pET-47b(+) were sequentially digested as described in methods section 2.2.2. Once both digests had been completed the DNA fragments were purified using a QIAquick PCR cleanup kit (Qiagen, Loughborough, UK). The concentration of DNA was calculated.

pET-47b(+) was selected with the goal of expressing His-tagged proteins in the cytoplasm without the signal peptide, hence the exclusion of the signal sequence by primer design. Ligation reactions were carried out at pET-47b(+):gene ratios: 1:1, 1:3, 1:6 and 1:10. Negative controls were included as described in section 2.2.5. When complete the reaction mixes were used to transform NEB Turbo Competent *E. coli* cells which were plated onto agar plates for antibiotic screening. Development of the method is included in the methods (section 2.2.5).

The two ligation ratios that proved most effective were 1:3 and 1:6. Colonies of *E. coli* containing BAPKO\_0422/pET-47b(+) were produced at both of these ratios, whilst colonies of BG0407/pET-47b(+) were produced at the ratio 1:3. Multiple colonies from each plate were selected and cultured overnight in 5 ml LB with kanamycin (as described in section 2.2.6). Potential constructs were extracted from the cultures using a QIAprep Spin Miniprep Kit (Qiagen, Loughborough, UK) and eluted in 50 µl nuclease free water. 2 µl of each potential construct digested using both restriction enzymes. The resulting DNA fragments were analysed using a 2 % agarose gel (Figure 4.8).



**Figure 4.8. Digested gene constructs.** L-R 1 kbp step ladder, BAPKO\_0422/pET-47b(+) constructs, BG0407/pET-47b(+) constructs. Bands at approximately 600-700 bp are inserted DNA (white box) were excised from the construct by digestion.

Once it was confirmed by electrophoresis that two fragments were present at approximate sizes (vector and inserted DNA) the BAPKO\_0422 and BG0407 constructs were sent for sequencing at Source Bioscience (Nottingham, UK). Sequencing was conducted using T7 primers in both the forward and reverse direction (Figure 4.9 & Figure 4.10).

0422 (1) gene with His-tag.

ATG GCA CATCACCACCACCACCAC TCCGCGGCT CTTGAAGTCCTCTTTCAGGGACCC GGGT  
 ACCAGGATCCGCAATCAAAAAGCAAACTATGGTTGAAGATGATTTTGACTTTGAGAACT  
 TCTTGAAAAAGAAGAGTCTGTGCGCCGTTTATTTGGCATAGGCTTTGGAATTGGATACCCA  
 CTTACAAACATTACAATATCTGTTGCGTATGTAGATATAGACCTCGGCTACGGAAGATTCTG  
 TTGGGCTTAAACCCAACAATTTTCATGCCCTATGTTGTAATGGGAATAGACCTTTTAATTAA  
 AGATGAAATACATAAAAACACTATGATTTCCGGCGGCATTGGAATGGGTGCAGATTGGTCA  
 AAAGGAAGTCCTGAAAAATCGAATGAAAACCTTGAAGGAGACGTAAATGAAGATCAACAAA  
 CATCTCTTGAAAATAGAATAGGGGTTGTAATAAGGCTACCTTTGGTAATAGAATACAGTTT  
 TCTTAAAAATATTGTAATTGGGTTTAAAGCTGTTGCTACTATTGGAACAACCTATGCTATTT  
 GGAAACCCAATGTCATTTGAAGGAGCTAGATTTAATTTCTTGGGCACAGGCTTTATAAAAA  
 TATATATATAA

**Figure 4.9. BAPKO\_0422 gene sequence obtained from construct.** Sequencing carried out at Source Bioscience (Nottingham, UK) using the T7 promoter in the 5'-3' direction. This sequencing data is much clearer than the sequencing of just the gene. No manual adjustments to the data were necessary. Highlighted in green is the start codon, turquoise is the His-tag, red denotes the HRV-3C cleavage site and grey the BAPKO\_0422 coding sequence.

The sequenced construct DNA for BAPKO\_0422 and BG0407 were translated using the TRANSLATE program on EXPASY. This found that the BG0407 gene was out of frame with the open reading frame of the plasmid (Figure 4.10); resulting from an error in primer design - an extra base pair was required between the restriction site and gene to bring the gene into the reading frame of the plasmid.

**BAPKO\_0422 FORWARD** MAHHHHHH SAA LEVLFQGP GYQDP QSKSKTMVEDDFDFEKL  
 EKEESVRRLFGIGFGIGYPLTNITISVAYVDIDLGYGRFVGLKPNNFMPYVVMGIDLLIKD  
 EIHKNMTISGGIGMGADWSKGSPEKSNEENLEGDVNEQQTSLENRIGVVIRLPLVIEYSFL  
 KNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI-

**BAPKO\_0422 REVERSE** MAHHHHHH SAA LEVLFQGP GYQDP QSKSKTMVEDDFDFEKL  
 EKEESVRRLFGIGFGIGYPLTNITISVAYVDIDLGYGGFVGLKPNNFMPYVVMGIDLLIKD  
 EIHKNMTISGGIGIGADWSKGSPEKSNEENLEGDVNEQQTSLENRIGVVIRLPLVIEYSFL  
 KNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI-

**BG0407 FORWARD** -SPLSGTRVPGSD QSKSKSMVEDDFDFEKL LAKEESVRRLFGIGFG  
 IGYPPLTNITISVPYVDIDLGYGGFVGLKPNNFMPYVAAGIDLLFKDEIHKNTMISGGIGIG  
 ADWSKGSPEKSNEKLEEEENEAAQQVNSLQNRIGVVIRLPLVIEYSFLKNIVIGFKAVATI  
 GTTMLLGGPMSFEGARFNFLGTGFIKIYI-

**Figure 4.10. Translation of sequenced constructs.** Top: BAPKO\_0422 sequenced in the T7 forward direction, middle: BAPKO\_0422 sequenced in the T7 reverse direction, bottom: BG0407 sequenced in the T7 forward direction. Highlighted in grey is the protein sequence, green is the start codon, the His-tag is turquoise and the HRV-3C cleavage site red. The His-tag and HRV-3C cleavage site are not present in the same reading frame as the BG0407 gene. The cloning primers were reanalysed and it was found that the forward BG0407 primer was incorrect, resulting in the gene being out of frame by one base from the reading frame of the plasmid.

The resulting protein sequence in frame with the His-tag was BLAST searched. The top hit was BAPKO\_0422 (named BAPKO\_0408 from this strain) (Figure 4.11).

tr	Q0SNA2 Q0SNA2_BORAP	Putative uncharacterized protein [BafPKo_0408] [Borrelia afzelii PKo]	201 AA
Score = 358 bits (920), Expect = 2e-97 Identities = 179/182 (98%), Positives = 179/182 (98%)			
Query: 1	QSKSKTMVEDDDFDFEKLLEKEESVRRLF	GIGFGIGYPLTNIISVYVDIDLGYGGFVGL	60
Sbjct: 20	QSKSKTMVEDDDFDFEKLLEKEESVRRLF	GIGFGIGYPLTNIISVYVDIDLGYGGFVGL	79
Query: 61	KPNNFMPYVVMGIDLLIKDEIHKNTMISGGIGIGADWSKGSPEKS	SNENLEGDVNEDQQTS	120
Sbjct: 80	KPNNFMPYVVMGIDLLIKDEIHKNTMISGGIGIGADWSKGSPEKS	SNENLEGDVNEDQQTS	139
Query: 121	LENRIGVVIRLPLVIEYSFLKNIVIGFKAVATIGTTMLFGNPMSFEGARFN	FLGTGFIKI	180
Sbjct: 140	LENRIGVVIRLPLVIEYSFLKNIVIGFKAVATIGTTMLFGNPMSFEGARFN	FLGTGFIKI	199
Query: 181	YI	182	
Sbjct: 200	YI	201	

**Figure 4.11. Top hit from the BLAST search of BAPKO\_0422.** Protein sequence used is that from Figure 4.10 minus the His-tag, cleavage site and extra residues before the start of the protein. Polymorphisms between the cloned wild-type sequence and the database sequence are highlighted in turquoise. Q0SNA2 was the third hit behind homologues E2JNC3 - BafACA1\_0410 and K4IYN5 – BafHLJ01\_0440 each with scores of 361 and E-values of 5e-98. BafPKo\_0408 is another identifier for BAPKO\_0422 from a different isolate.

The gene amplified and cloned into pET-47b(+) was determined as BAPKO\_0422 with three amino-acid polymorphisms compared to the sequence in the database. Two polymorphisms seen in the sequenced gene (Figure 4.7) are not present in the construct gene. This is either due to more accurate sequencing data from construct DNA or the presence of different strains of wild-type *B. afzelii* in the tick and subsequent amplification of genes with polymorphisms.

## 4.3 Recombinant Expression of BAPKO\_0422

### 4.3.1 Codon usage

*E. coli* K12 codon usage was compared to the sequence of BAPKO\_0422 to determine if any rare codons used by *B. afzelii* would prevent optimum expression. The genetic sequence of amplified BAPKO\_0422 including His-tag and HRV-3C cleavage site was compared against *E. coli* K12 codon usage using the web based server - graphical codon usage analyser (gcua). Five codons present in the gene (CTC, ATA, AGA, AGG and CTA) are rarely used in *E. coli* K12 (Figure 4.12).



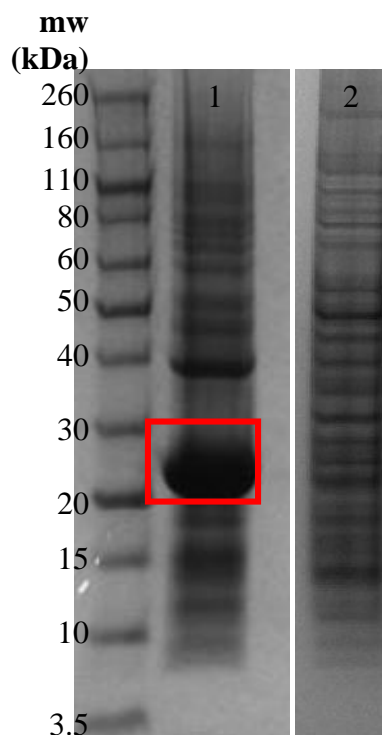


Due to the presence of rare codons in the gene, *E. coli* Rosetta (DE3) (Merck) were used for expression. Rosetta cells contain tRNAs for six rare codons on a chloramphenicol resistant plasmid to enable translation of almost all DNA sequences that contains rare codons. The six rare tRNAs accounted for are AGG, AGA, AUA, CUA, CCC, GGA, five of which feature in the gene sequence of BAPKO\_0422.

### **4.3.2 Autoinduction**

The ultimate goal was the production of an autoinduction protocol followed by purification and refolding of large volume of protein. Autoinduction was the preferred method of overexpression due to the potential to produce large quantities of protein. Initially however, small scale expression trials took place using 50 ml cultures of BAPKO\_0422 in LB and IPTG induction to test the Rosetta cells and the purification technique. Expression induced by 0.1 mM IPTG produced very little protein whereas expression induced by 1 mM IPTG produced larger quantities. Expression was accomplished using the protocol as described in section 2.2.7.1. and confirmed by SDS-PAGE and western blot (not shown).

Expression by autoinduction was conducted in a total volume of 400 ml according to the protocol detailed in section 2.2.7.2. The first attempt was unsuccessful. This was later determined to be due to the addition of the 5052 (glucose and lactose) component prior to autoclaving, resulting in caramelisation of the sugars. To counter this, 5052 was filter sterilised (0.45  $\mu$ M) and added after autoclaving of the base media. NPS and  $\text{MgSO}_4$  were also filtered and added after autoclaving to prevent compound degradation. The second expression was also unsuccessful, so the shaking speed was increased to 200 rpm and the foil seal removed to increase aeration of the culture. These changes resulted in a large quantity of protein being expressed, as confirmed by SDS-PAGE (Figure 4.13).



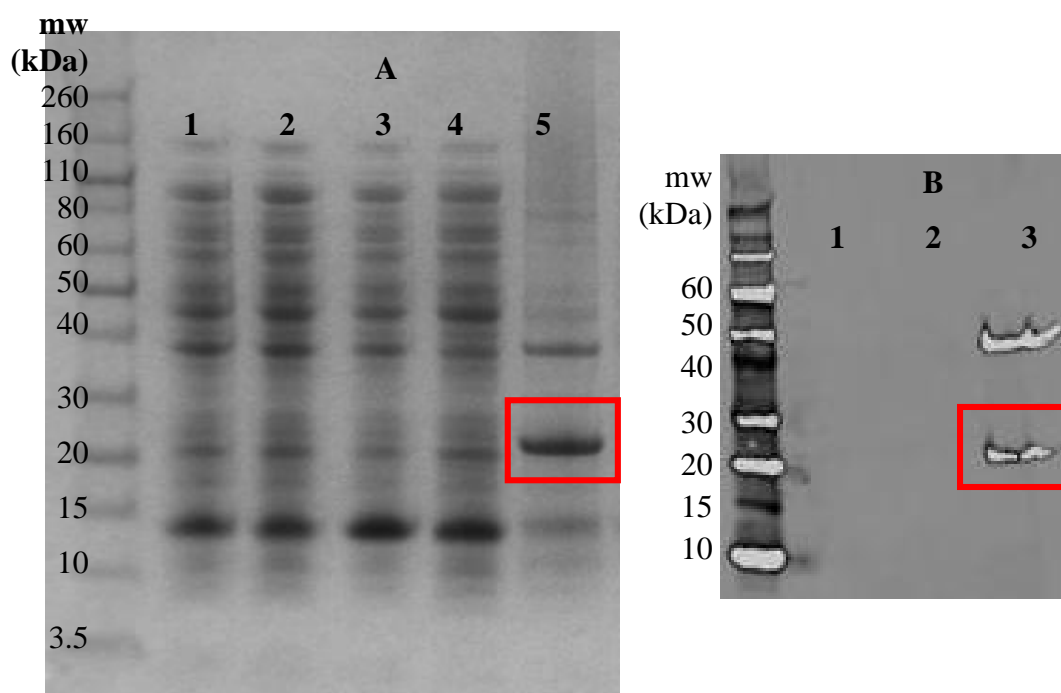
**Figure 4.13. Expression of BAPKO\_0422.** Whole cell lysates were run on an SDS-PAGE to determine expression of BAPKO\_0422. Lane 1 - The large band at approximately 23 kDa (red box) is expressed BAPKO\_0422. Lane 2 - Expression of *E. coli* with no expression construct.

## 4.4 Purification and refolding of BAPKO\_0422

### 4.4.1 Purification of inclusion bodies

Development of the lysis buffer and insoluble protein separation protocols was a substantial process; initially the harvested cell pellet was resuspended in 1x PBS and lysed by sonication for three minutes before washing. From visual inspection it appeared that lysis was incomplete. Varying sonication pulse lengths were tried before the initial resuspension buffer was switched from PBS to 0.3 M NaCl, 50 mM Tris-HCl at pH 8 and one hour incubation on ice with 0.4 mg/ml lysozyme was added prior to the sonication step. This modified protocol proved adequate for lysing the volume of cells present in < 50 ml IPTG induced cultures. Autoinduction cultures however contained much larger volumes of cells, which were not completely lysed by the protocol for IPTG induction, hence the concentration of lysozyme was increased to 1 mg/ml and sonication time increased to two minutes total pulse. Detergents such as Triton X-100 are often added to buffers to assist in the lysis of cell membranes. Initially this was not considered an option as many detergents are used to solubilise membrane proteins from inclusion bodies (Seddon *et al.*, 2004) and as solubilising the protein would mean refolding, it was preferable to leave the protein in inclusion bodies until purification. A trial was conducted using 0.1 % (v/v) Triton X-100 added to the lysis

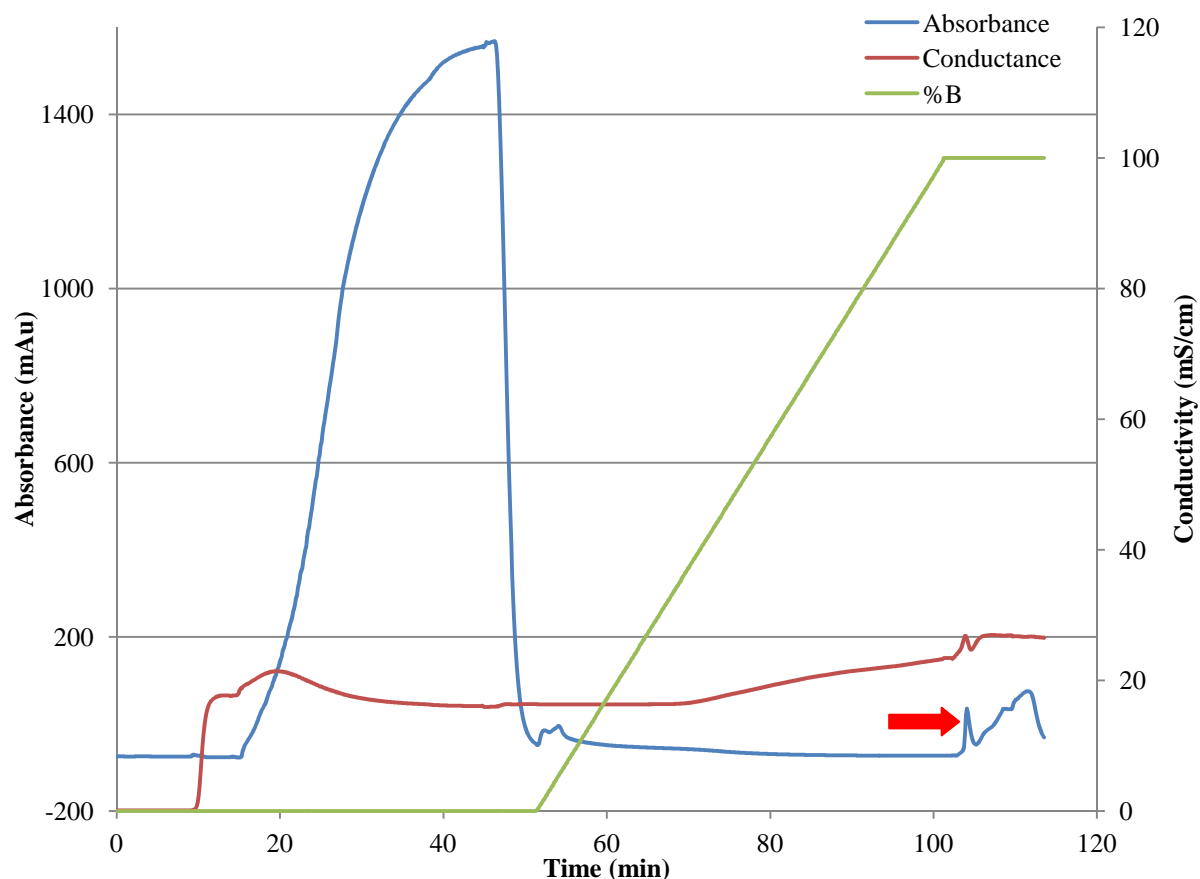
buffer as Triton X-100 is not generally used for membrane protein refolding. Lysis was much more efficient with the addition of Triton X-100 and SDS-PAGE analysis confirmed that the protein was still present in inclusion bodies. The inclusion bodies were solubilised in FPLC buffer A (Table 2.2) after three salt washes to remove soluble proteins (described in section 2.2.9) and the soluble and insoluble fractions analysed by SDS-PAGE and Western Blot (Figure 4.14).



**Figure 4.14. Separation of inclusion bodies. A) SDS-PAGE. B) Western blot.** Cells were harvested from an autoinduction culture, lysed and the soluble and insoluble fractions separated as previously described. **(A)** Protein standard ladder (Novagen), 1,2,3 - initial salt wash fractions, 4 - secondary salt wash fraction, 5 - insoluble fraction solubilised in 8 M urea. **(B)** Protein standard ladder (Novagen), 1 - initial salt wash fraction, 2 - secondary salt wash fraction, 3 - insoluble fraction solubilised in 8 M urea. As the initial salt wash fractions were all very similar, only one was run on the western blot. Highlighted in red is a band of approximately the expected size for BAPKO\_0422. No BAPKO\_0422 was present in either salt washes (soluble fractions). The gel used was a pre-cast 4 – 12 % (w/v) Bis/Tris run as described (section 2.2.8). A large number of proteins have been removed by the salt wash steps although several remained in the insoluble fraction and were present on the SDS-PAGE as faint bands. The secondary band at approximately 50 kDa that produced a positive result on the Western blot is either caused by an *E. coli* protein bound to BAPKO\_0422 or a BAPKO\_0422 dimer.

#### 4.4.2 IMAC purification & refolding

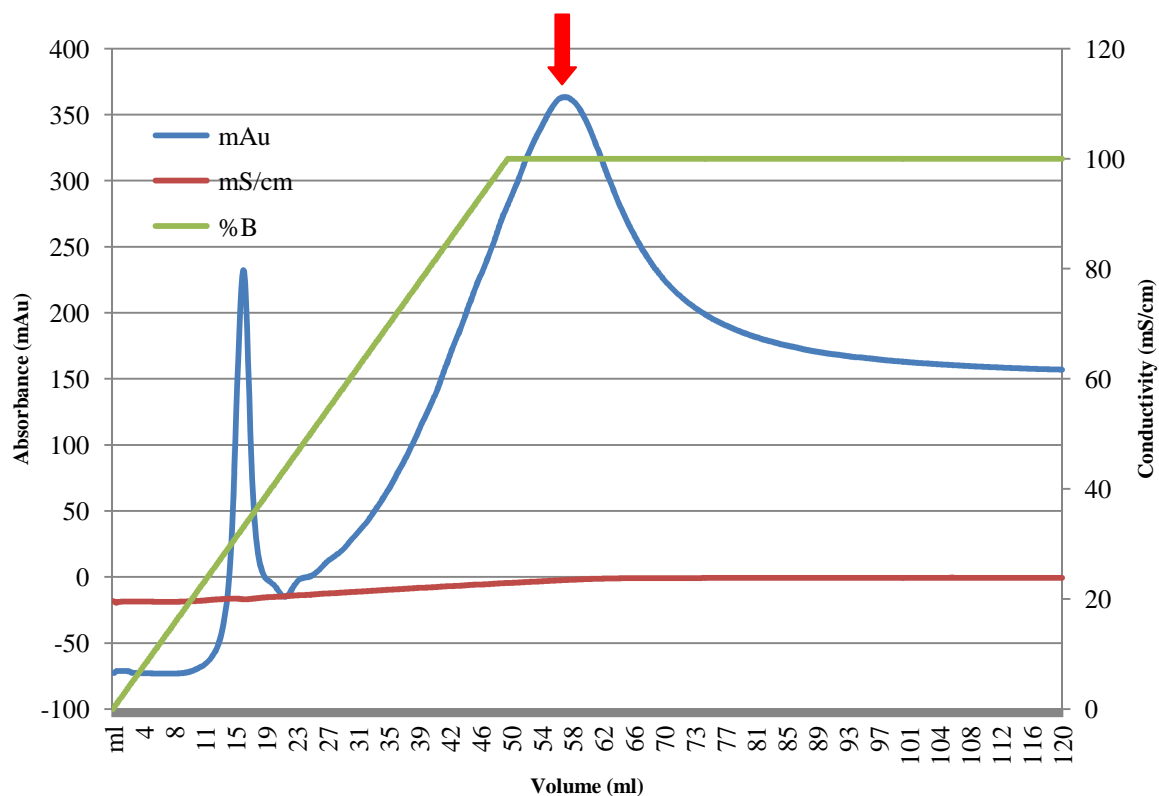
Cells from expression cultures were lysed and protein harvested from inclusion bodies as described in section 2.2.9. Initial buffers for resolubilisation of inclusion bodies and IMAC purification contained 8 M urea instead of 6 M GuHCl as this was the most common denaturant used in the literature. However, this resulted in very low binding of protein to the column with a large amount of protein remaining in the flow through (Figure 4.15).



**Figure 4.15. IMAC Chromatogram of purification of BAPKO\_0422 in inclusion bodies solubilised with 8 M urea.** Flow rate used was 1 ml/min<sup>-1</sup>. Sample was loaded in 8 M urea, 0.3 M NaCl, 30 mM Tris-HCl, pH 8 and run with this buffer until the baseline returned to 0. The large peak from 20 to approximately 50 minutes is the loading peak of protein that did not bind to the column. Protein was eluted by buffer containing 0.3 M imidazole, 8 M urea, 0.3 M NaCl, 30 mM Tris-HCl, (pH 8). The small peak at around 105 minutes (red arrow) is the eluted protein, followed by the peak caused by absorbance of light by imidazole. Only a minute fraction of protein bound to the column, hence protocol optimisation was required.

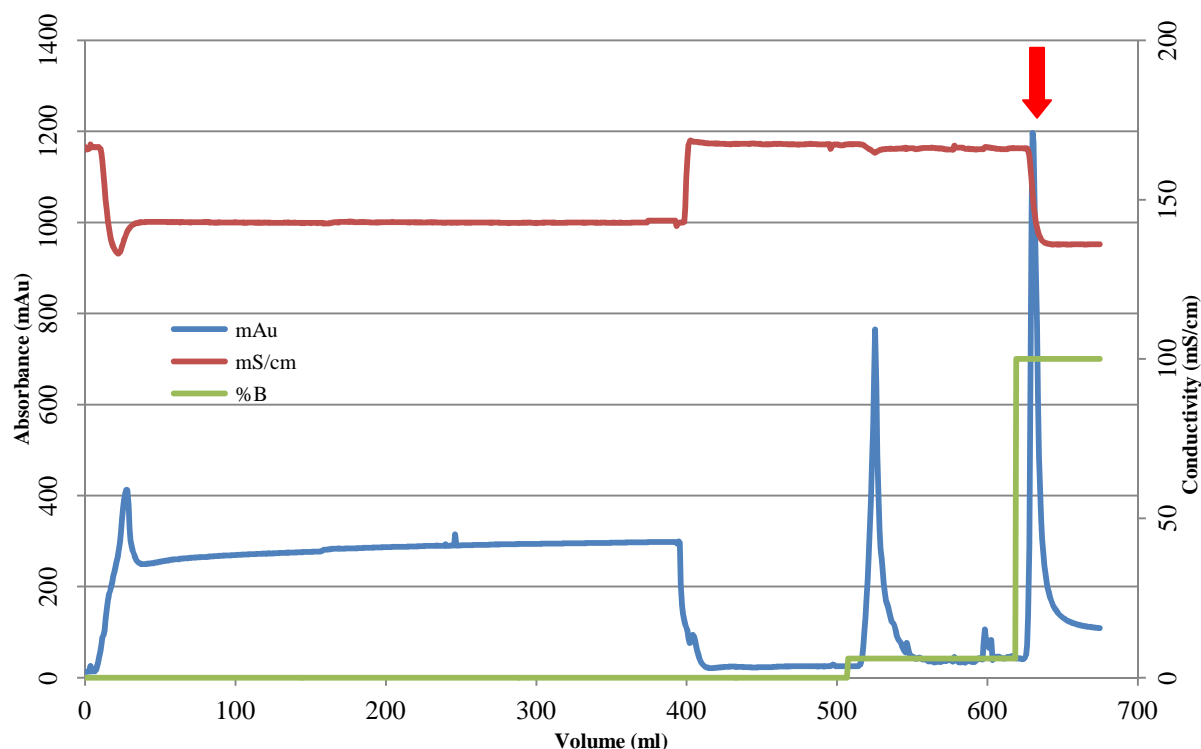
In an attempt to increase the binding capacity, the pH of all buffers was varied (7, 8 and 9), and the NaCl concentration also varied (0.1, 0.3, 0.5 and 1 M). This included the buffer in which the inclusion bodies were resuspended. None of these changes or combinations of changes made any noticeable difference to the binding. The column was also cleaned with 0.5

M NaOH, stripped of nickel using 10 mM EDTA and recharged with freshly made and filtered (0.45  $\mu$ m) 100 mM NiSO<sub>4</sub>, but again no increase in binding was found. The denaturant was switched from 8 M urea to 6 M GuHCl at pH 8, this showed an increase in protein binding (Figure 4.16).



**Figure 4.16. IMAC Chromatogram of purification of BAPKO\_0422 in inclusion bodies solubilised with 6 M GuHCl.** This chromatogram shows elution of the peak once the denaturant had been switched to GuHCl. When compared to the previous chromatogram the peak eluted had a higher absorbance and was wider (red arrow), indicating that more protein had bound to the column. The peak at approximately 15 ml was found to be non-specifically bound protein by SDS-PAGE, and the large peak to be pure BAPKO\_0422 (data not shown). Both conductivity and % buffer B are plotted on the secondary axis.

After this improvement, the pH and NaCl concentration were altered, but again to no avail. After searching the literature it was observed that glycerol was added to the sample and loading buffers to reduce hydrophobic interactions (Bane *et al.*, 2007). Glycerol was added to the sample and the 6 M GuHCl loading buffer (FPLC buffer A) to give a total 10 % (v/v), which resulted in an increase in protein binding (Figure 4.17).



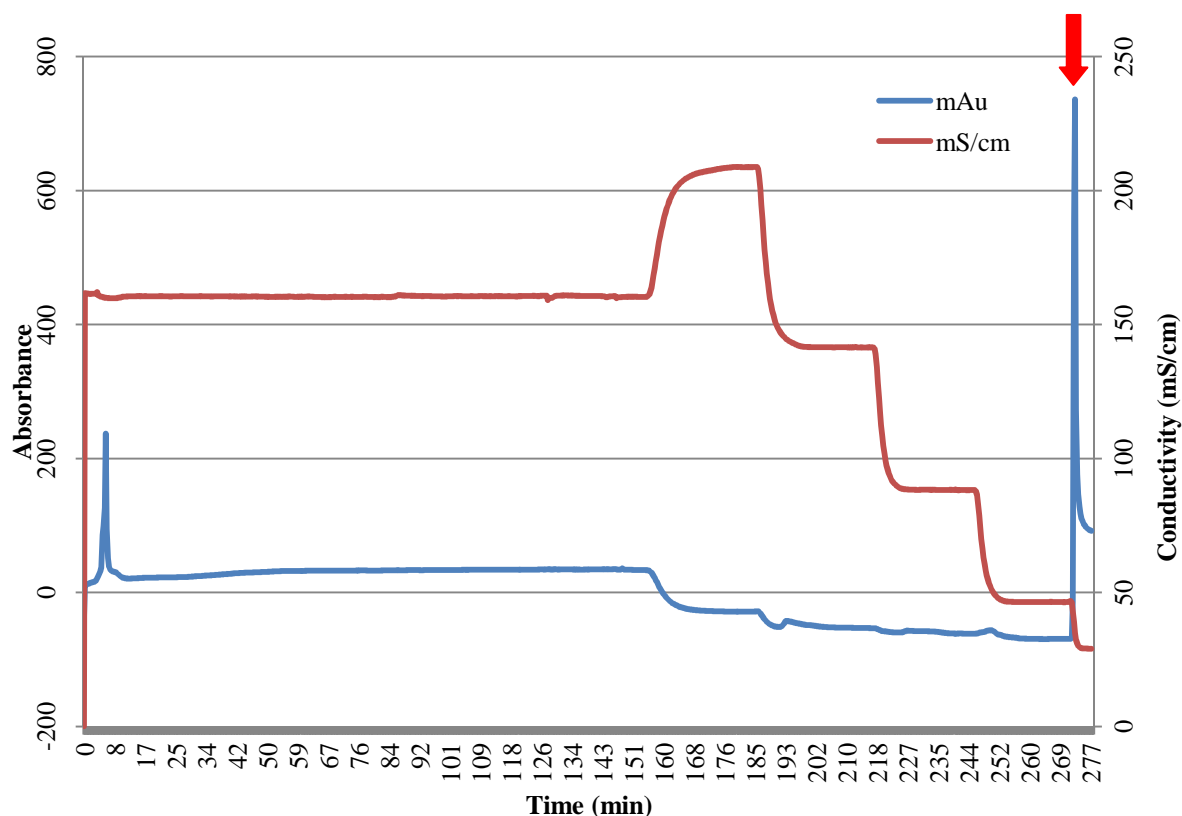
**Figure 4.17. IMAC Chromatogram of purification of BAPKO\_0422 in inclusion bodies solubilised with 6 M GuHCl with the addition of 10 % (v/v) glycerol.** The loading stage (0-400 ml) is longer than usual due to this being a sample that did not bind on a previous purification attempt. The peak at approximately 530 ml is due to weakly bound proteins eluting. The peak at approximately 640 ml (red arrow) is bound BAPKO\_0422 being eluted by the addition of an imidazole-containing buffer (6 M GuHCl, 0.3 M NaCl, 30 mM Tris, pH 8). Both conductivity and % buffer B are plotted on the secondary axis.

The final attempt at increasing binding affinity was the addition of 1 mM EDTA to the lysis buffer. As the *E. coli* were being cultured in LB and autoinduction media, there was a large number of metal ions present for growth. It was speculated that these metal ions could be strongly binding the His-tag, preventing the His-tag binding to the nickel on the column and so the chelating agent EDTA was added. EDTA is used for stripping the nickel atoms from the column, it was important that the EDTA never reached the column, hence its presence in the lysis buffer. The EDTA was removed along with soluble proteins before the sample reached the column, preventing the nickel being stripped from the column. Addition of EDTA produced a minor increase in binding affinity of BAPKO\_0422.

Development of the refolding protocol proceeded in conjunction with attempts to increase the binding affinity for the column. The small yield that was obtained was shown to be pure by SDS-PAGE (Figure 4.20) and so on-column refolding was attempted (Figure 4.18).

The  $\alpha$ -barrel domain of *E. coli* OmpA has been shown to spontaneously refold in detergent at room temperature (this was also trialled with BAPKO\_0422 in this research and estimated to be true by SEC), however it was decided that a one-step purification & on-column refold would be preferable to minimise purification steps. The on-column refolding protocol for BAPKO\_0422 protein was adapted from (Ito *et al.*, 2003), which used a gradual decrease in denaturant to achieve on-column refolding. Initially, three refold buffers were planned (Table 2.2), this would leave the protein in 1 M GuHCl before gradient elution which would further dilute the denaturant. A fourth RF buffer (the same composition as SEC buffer, Table 2.2) was added to completely eliminate GuHCl from the sample. Initially upon the exchange of one refold buffer to another, the column was left for half hour to equilibrate. Experimentation with equilibration times found that this equilibration step was not necessary for proper refolding and so removed, speeding up the process. During the refold process, no protein eluted from the column and a large peak was detected upon the addition of the imidazole-containing elution buffer (Figure 4.18).

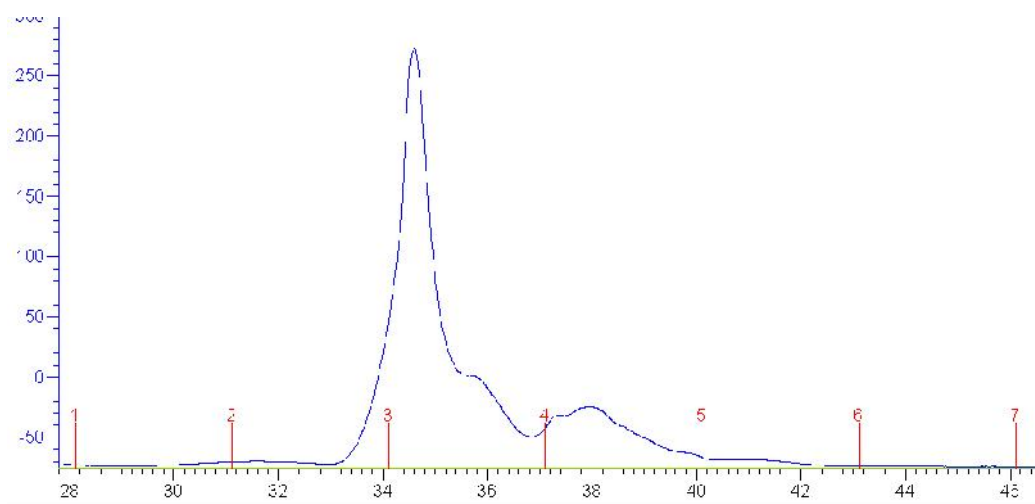




**Figure 4.18. IMAC Purification and on-column refold of BAPKO\_0422.** A small volume of previously purified BAPKO\_0422 was loaded onto the column and refolded. The change in refolding buffers is represented in the conductivity. At approximately 160 minutes the conductivity peaks, this is due to washing with buffer RF1 (4 M GuHCl, 0.3 M NaCl, 30 mM Tris, pH 8). Each decrease and plateau that follows is the exchanging of buffer. The GuHCl component decreased with each new buffer (6 M, 4 M, 2 M, 1 M and 0 M) The NaCl and Tris concentrations remained constant and 0.1 % (w/v) LDAO was included with the 4 M GuHCl buffer and all those used thereafter. BAPKO\_0422 elution was a result of the introduction of an imidazole-containing elution buffer (0.1 % (w/v) LDAO, 0.5 M imidazole, 0.5 M NaCl, 50 mM Tris-HCl, pH 8).

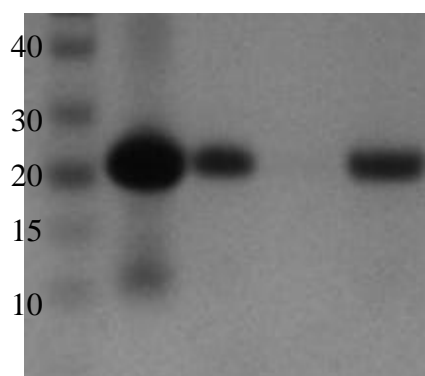
#### 4.4.3 SEC Purification

The column was prepared and calibrated as described in section 2.2.11. BAPKO\_0422 was concentrated to a volume of 1 ml using a 10,000 mwco before SEC. SEC was conducted using a flow rate of 0.3 ml/min with fractions collected every 3 ml. Fraction collection began after approximately 5 ml of buffer had run through the column. BAPKO\_0422 eluted at approximately 11 ml (Figure 4.19), which equates to a molecular weight of approximately 40 kDa. As has been noted, SEC cannot be used to accurately determine the mw of a membrane protein due to the binding of an unknown number of detergent molecules. However the expected mass of BAPKO\_0422 (~23 kDa) added to the approximate mass of an LDAO micelle (~17 kDa) results in a total mass of ~40 kDa, equivalent to that calculated from the elution volume using the calibration curve (section 2.2.11).



**Figure 4.19. SEC Chromatogram of BAPKO\_0422.** Screenshot of SEC chromatogram from PrimeView software shows a total elution on the X-axis. The red dashes are fraction numbers. 0.3 ml BAPKO\_0422 in 0.1 % (w/v) LDAO, 0.3 M NaCl, 30 mM Tris (pH 8) at a concentration of approximately 5 mg/ml was loaded. The clear peak at approximately 34.5 ml total volume was shown to be pure BAPKO-0422 by SDS-PAGE. Based on the elution volume of 11 ml and the calibration curve (Figure 2.5) this equates to a molecular weight of approximately 40 kDa.

Analysis of the protein peak was conducted by SDS-PAGE (Figure 4.20) and the protein found to be pure.



**Figure 4.20. SDS-PAGE of purified BAPKO\_0422.** L-R: Sharp unstained protein standard (Novex), insoluble protein fraction, protein purified by IMAC, empty well, protein purified by SEC. The insoluble protein fraction had a small number of proteins present besides BAPKO\_0422. These were all removed by IMAC. In addition there appears to be no anomalous migration of the protein despite being boiled in SDS. BAPKO\_0422 showed minimal absorbance at 320 nm, indicating no protein aggregation.

## 4.5 Discussion

The use of nested PCR (nPCR) was important for specific amplification of genes from the samples provided by Dr. Gabi Margos. As these DNA samples are extracted from whole ticks, *Borrelial* DNA was present in very small concentrations compared to tick DNA. As described in section 4.1 amplification of genes with inner amplification primers (targeting only the gene) was unsuccessful, therefore nested PCR using both primer pairs was necessary to achieve the sensitivity required. During amplification by nested PCR there were several instances of non-specific amplification (Figure 4.4), this was not unexpected as the high concentration of tick DNA and the relatively low  $T_m$  of primers used results in a high probability of non-specific primer binding. The main amplification was of the desired genes, which were excised from the gel where necessary and purified. An alternative to using nPCR would have been to use pure *Borrelial* DNA such as that provided by Dr Volke Fingerle, however this DNA was not obtained until after the genes had successfully been amplified.

A BLAST search of the translated sequence of amplified BAPKO\_0422 shows 97 % identity with that of BAPKO\_0422 from *B. afzelii* PKo. Some residues differ between the two sequences, such polymorphisms are expected as the DNA was extracted from different isolates. The region towards the C-terminus shows many unassigned residues (denoted as X) as the sequencing become unreliable towards the end, hence the sequencing was carried out in both directions.

Restriction digestion of the DNA fragments and plasmid vector proved problematic. Control digestions found that both dilute (1 in 5) and neat *HindIII* fully digested the vector, whereas neither neat or dilute *BamHI* accomplished complete digestion under recommended conditions. A doubling of the BSA concentration in conjunction with using neat enzyme stock in the *BamHI* digest resulted in almost complete digestion. Re-ligation of singly digested vector was conducted as a positive control, this required a doubling of ATP and T4 DNA ligase concentration to be successful. Ligation of doubly digested insert to vector also needed optimising. Various ratios were used as described in the methods (section 4.2). The ligation temperature was varied: 15 °C, 20 °C, 23 °C, 33 °C and 37 °C, as was the incubation times: 1, 1.5 and 2 hours. The various temperatures tried were suggested by the manufacturer's instructions and from troubleshooting forums. The conditions under which ligation as successful (albeit with a low success rate) were: incubation at 23 °C for 1 hour

with the addition of shaking at 120 rpm using doubled ATP and T4 ligation concentration. Shaking was added in an attempt to encourage more movement in the reaction and increase interactions between molecules. The addition of more base pairs to the 5' end of the each restriction primer (ideally a total minimum of six bases should be included) would result in the improved efficiency of the restriction enzyme as many restriction enzymes inefficiently cleave restriction sequences near the ends of a DNA fragment, in particular those generated by PCR amplification (Sambrook *et al.*, 2000). Alternatively a different cloning approach that did not require the use of restriction enzymes such as blunt-end or TA cloning would be used as this would remove the need for restriction digestion entirely. Blunt end cloning however may result in genes inserted backwards (which would require sequencing to confirm), and both techniques have the potential to insert multiple copies of the gene.

Analysis of the protein sequences of the constructs translated from sequencing data (Figure 4.10) highlighted an error in the forward primer for addition of restriction sites to BG0407; the protein sequence was not in the same reading frame as the His-tag, meaning that the protein was not expressed. The BAPKO\_0422 construct was produced in frame with the His-tag. A BLAST search of the protein sequence to be expressed found an excellent match for BAPKO\_0422, with three polymorphisms. The first two polymorphisms I42T and P46A are predicted to be on the inside of a  $\alpha$ -strand; both polymorphisms result in a smaller side-chain protruding into the centre of the barrel, potentially altering the inner H-bonding networks. The third predicted polymorphism F77I is predicted to be present at a residue forming part of the aromatic girdle, substituting a phenylalanine for an isoleucine; this is not ideal for the girdle, however both OmpA and OmpX from *E. coli* have an isoleucine present in the girdle, hence this is not an unreasonable substitution. It is also possible that the prediction of  $\alpha$ -strand locations is not entirely correct.

The codon usage by BAPKO\_0422 was checked for rare codons (Figure 4.12) and it was found that several codons not used by *E. coli* K12 were present in the gene sequence. Hence *E. coli* Rosetta (DE3) cells (Merck) were used. Induction of expression by IPTG was successful using standard lab protocols, however expression by autoinduction required an increase in shaking from 120 rpm to 200 rpm and removal of the foil seal; an increased incubation time and amplified proliferation of *E. coli* requires much more sample aeration. Standard *E. coli* BL21 cells used as a negative control produced no expression under the same conditions, validating the necessity of Rosetta cells.

Purification of BAPKO\_0422 by IMAC suffered primarily from low binding efficiency of the protein to the column. A variety of alterations were made to the buffer composition and protocol in an effort to increase binding. Increasing or decreasing the pH and the concentration of NaCl of the loading buffer either individually or in tandem had no effect on binding. Switching the denaturant from urea to GuHCl had the effect of increasing protein binding slightly. The addition of 10 % (v/v) glycerol inhibit non-specific interactions also resulted in increased binding, as did the inclusion of 1 mM EDTA in the lysis buffer. This was done as the autoinduction media contained metal ions and EDTA is a strong chelating agent. The EDTA would therefore strip any metals that had bound the His-tag, which would block binding to the Ni-NTA column. Columns from Qiagen and GE Healthcare were used as bought and also stripped and recharged with freshly made NiSO<sub>4</sub>; the brand of column and state of nickel had no discernible effect on binding affinity. BAPKO\_0422 was eluted through the use of 0.3 M imidazole, although 1 M imidazole was used to determine if there was any strongly binding protein that was not eluting under the influence of 0.3 mM imidazole, hence creating the appearance of weakly binding protein. This was not the case.

Optimisation suggested that either the His-tag was not properly exposed due to aggregation formed by hydrophobic interactions, or that the His-tag was being hindered somehow by being present at the N-terminus of the protein. It is more likely that the protein was not in an optimal position at the N-terminus as hydrophobic interactions should have been disrupted by the addition of glycerol. A different purification method such as ion exchange may have more success due to the large positive charge present (His-tag) or perhaps a different tag such as a SUMO-tag could be used.

It was possible to refold BAPKO\_0422 in LDAO on the benchtop as described (section 2.2.10.2). However, a protocol for refolding of the protein on the column was conducted to increase purification efficiency. This was achieved using a system of buffers to lower the denaturant concentration in the presence of an introduced detergent. Initially the each refolding buffers were equilibrated on the column and left to sit for 30 minutes to allow refolding to occur. Optimisation of the method found that this was unnecessary and that the buffers could be loaded and once equilibrated the next buffer loaded. In the future the next optimisation step would be to trial a gradient from 6 M to 0 M GuHCl using just two buffers; this would save preparation time and provide further automation for the process. The

successful refold of protein by both methods was confirmed (after SEC) by Circular Dichroism (chapter 5).

Size exclusion chromatography is a technique used to separate proteins based on the molecular weight and overall size and shape. The elution volume can be used to calculate the molecular weight of the protein based on a calibration curve of proteins of known size and shape. In the case of BAPKO\_0422, SEC was used to obtain an estimation of sample purity and not an accurate approximation of molecular weight. As BAPKO\_0422 is a TM protein and requires detergent to 'protect' the hydrophobic core, a large unknown amount of detergent binds the hydrophobic region in an unknown configuration (section 2.2.16.3, Figure 2.7). Thus the overall size, shape and molecular weight of the protein-detergent complex is unknown, preventing accurate calculation of the molecular mass or oligomeric state. A sharp peak was present, suggesting that at least a portion the protein was pure; however the peak was followed by several smaller peaks. This is potentially a result of differing numbers of detergent molecule binding the hydrophobic region, with a large number forming something similar to a complete micelle and the later eluting molecules having fewer detergent molecules bound. There should be no molecules smaller than ~10 kDa present as the sample was concentrated by centrifugation using a 10,000 Da mwco concentrator. A very rough approximation of molecular weight was calculated using the elution volume of the peak and assuming that a full LDAO detergent micelle formed around the hydrophobic core. From the calibration curve (section 2.2.11, Figure 2.5) it was approximated that the protein and LDAO micelle were present in a 1:1 ratio. Analysis of the absorbance of BAPKO\_0422 at 320 nm showed very little absorbance, suggesting minimal aggregation.

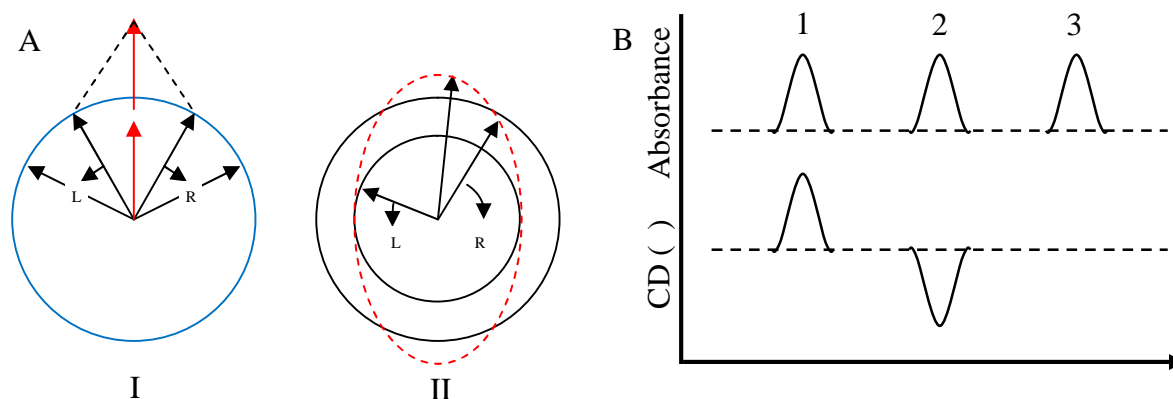
## **Chapter 5. Structural and functional studies of BAPKO\_0422**

### **5.1 Circular dichroism**

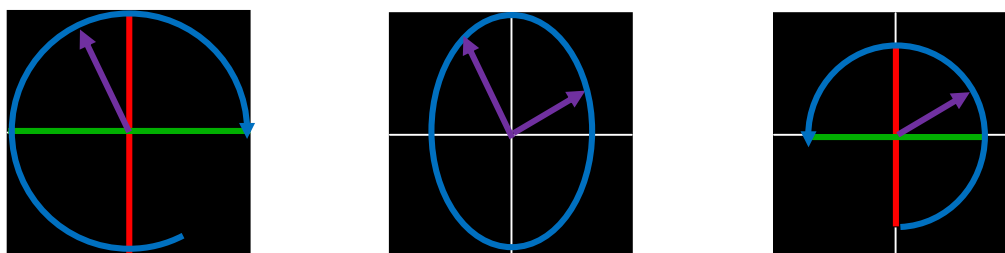
Protein circular dichroism (CD) is the measure of the difference in absorbance of right and left handed circularly polarised light over a range of wavelengths by an optically active sample, and the use of this data to calculate the percentage composition of protein secondary structure (Kelly *et al.*, 2005). CD is an important technique in determining the secondary structure of proteins as it can be carried out in a variety of buffers under denaturing conditions, allowing for observation of conformational changes (due to ligand binding, temperature or the presence of denaturants for example) and differences in structure between

mutants (Greenfield, 2006, Kelly *et al.*, 2005). Whilst CD is a low resolution technique compared to X-Ray crystallography and NMR spectroscopy, the ability to calculate the overall percentage composition of protein secondary structure using CD is a useful technique to aid the determination of high resolution structures by providing preliminary data with regards to the structure and stability (Kelly *et al.*, 2005).

The paths of left-handed (L) and right-handed (R) polarised light trace out circles (hence circular polarised light), and are 90 ° out of phase with each other (Greenfield, 2006). Assuming that the magnitudes of L and R are equal to each other, when viewed from the front, this wave appears as two equal length circular vectors, one left handed and one right handed (Figure 5.1 A (I), blue circle). The addition of the vectors produces planar polarised light (red arrows, Figure 5.1 A (I)). This occurs when L and R are absorbed equally by samples, or not at all (achiral molecules). This does not produce a CD signal. However, unequal absorption of L and R by optically active samples results in the plane of the light wave being rotated upon addition of the vectors creates an ‘elliptical’ polarisation (red ellipse, Figure 5.1 A (II) & Figure 5.2) (Kelly *et al.*, 2005). This is the origin of the CD effect (Greenfield, 2006).



**Figure 5.1. The CD effect.** A) Left and Right polarised light in the plane of polarised radiation. (I) – both L and R have the same polarisation magnitude, which when combined generate planar polarised light (red arrows). (II) L and R are polarised to different degrees, producing an elliptically polarised signal (red dotted line). B) Absorption and CD spectra, 1 – L absorbed more than R, 2 – R absorbed more than L, 3 – an achiral molecule, from which no CD signal is observed. Diagram based on (Kelly *et al.*, 2005).



**Figure 5.2. Addition of L and R polarised light of different magnitudes.** A) Right-handed polarised light, B) R absorbed more than L, addition of the two vectors results in elliptically polarised light and the CD signal, C) Left-handed polarised light.

The generated CD signal is reported in many units, the most common of which are; differential absorbance (  $\Delta A$  ), molar circular dichroism (  $\Delta \epsilon$  ), degrees of ellipticity (  $\theta$  ), and molar ellipticity [  $\theta$  ]. The CD signal is calculated using the equation:

$$\Delta A = A_L - A_R$$

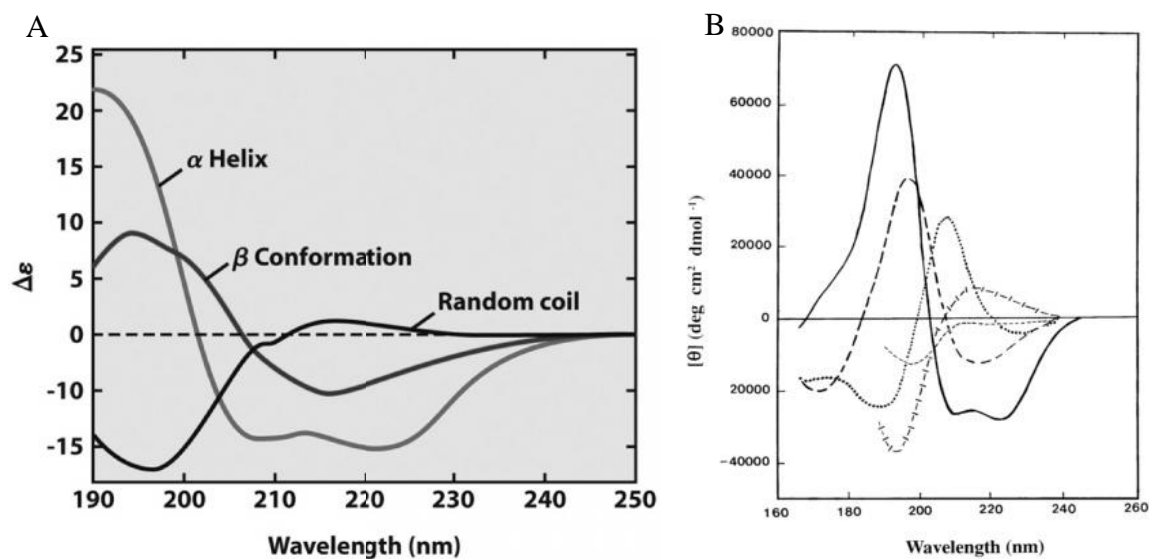
**Equation 3. Calculation of  $\Delta A$ .**  $A_L$  and  $A_R$  are the absorption of left and right handed polarised light respectively. For achiral molecules,  $\Delta A = 0$  as the absorption of R and L will both be 0, this is not applicable to small chemical molecules however.

Degrees of ellipticity is defined as the tangent of the ratio of the minor to major elliptical axis (Greenfield, 2006). The magnitude of  $\theta$  varies depending upon the concentration of protein in the sample and path length. Hence molar ellipticity [  $\theta$  ] is more commonly used; molar ellipticity is the CD signal corrected for concentration and path length (  $\theta / (\text{molar concentration} \times \text{path length})$  ) and has units of degrees  $\text{cm}^{-2} \text{dmol}^{-1}$  (James and Wallace, 2009). The CD spectrum of a sample is obtained by CD signal as a function of wavelength (Kelly *et al.*, 2005).

Different structural elements of proteins have different characteristic CD spectra (Figure 5.3).

$\alpha$ -helices have negative peaks at 208 and 222 nm and a positive peak at 193 nm,  $\beta$ -sheets have a negative peak at 218 nm and a positive peak at 195 nm, proteins that are disordered give a very low ellipticity signal above 210 nm and have a negative peak at approximately 195 nm (Greenfield, 2006). The CD signals characteristic of each type of secondary structure arise due to differential absorption by the peptide bond (amide chromophores) between residues that are either chiral or present in chiral environments created by different types of ordered folding (Kelly *et al.*, 2005). The peptide bond has two electron transitions that contribute to CD signal, one at  $\sim 220$  nm and the second at  $\sim 190$  nm.

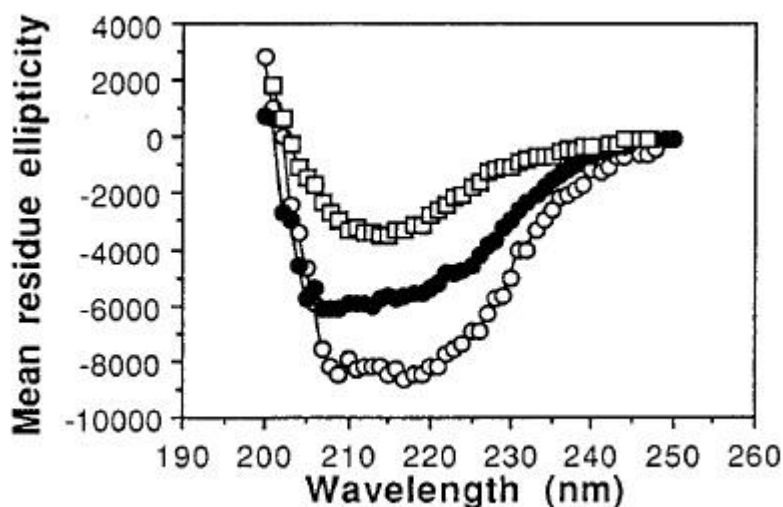




**Figure 5.3. Example CD spectra.** An example of far UV CD spectra of various secondary structures. **A)** a basic representation of three main secondary structure spectra (Nelson and Cox, 2008) reported as molar circular dichroism. **B)** More detailed secondary structure CD spectra (Kelly *et al.*, 2005) reported as molar ellipticity. Solid line;  $\alpha$ -helix, long dashed line; anti-parallel  $\beta$ -sheet, dotted line; type I  $\beta$ -turn; cross and dashed line, extended  $3_1$ -helix; short dashed line, random coil.

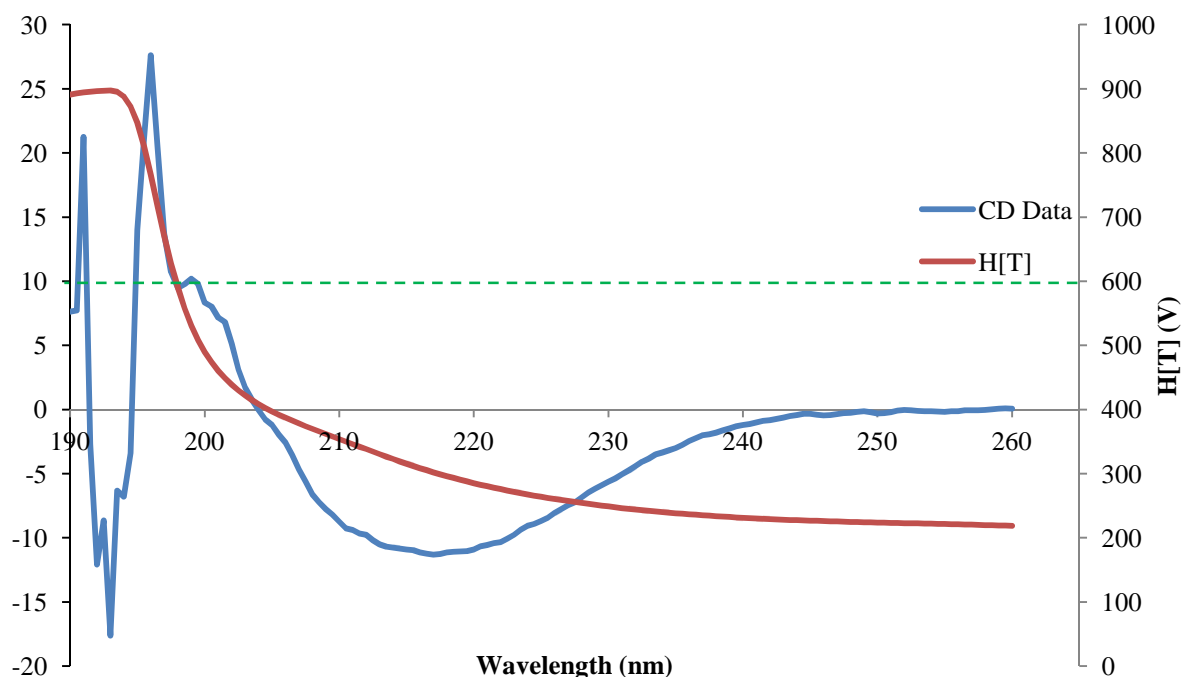
### 5.1.1 Circular dichroism of BAPKO\_0422

The N-terminal domain of *E. coli* OmpA has previously been analysed (Figure 5.4 - white squares). This provided a known reference spectra for a small TM  $\beta$ -barrel.



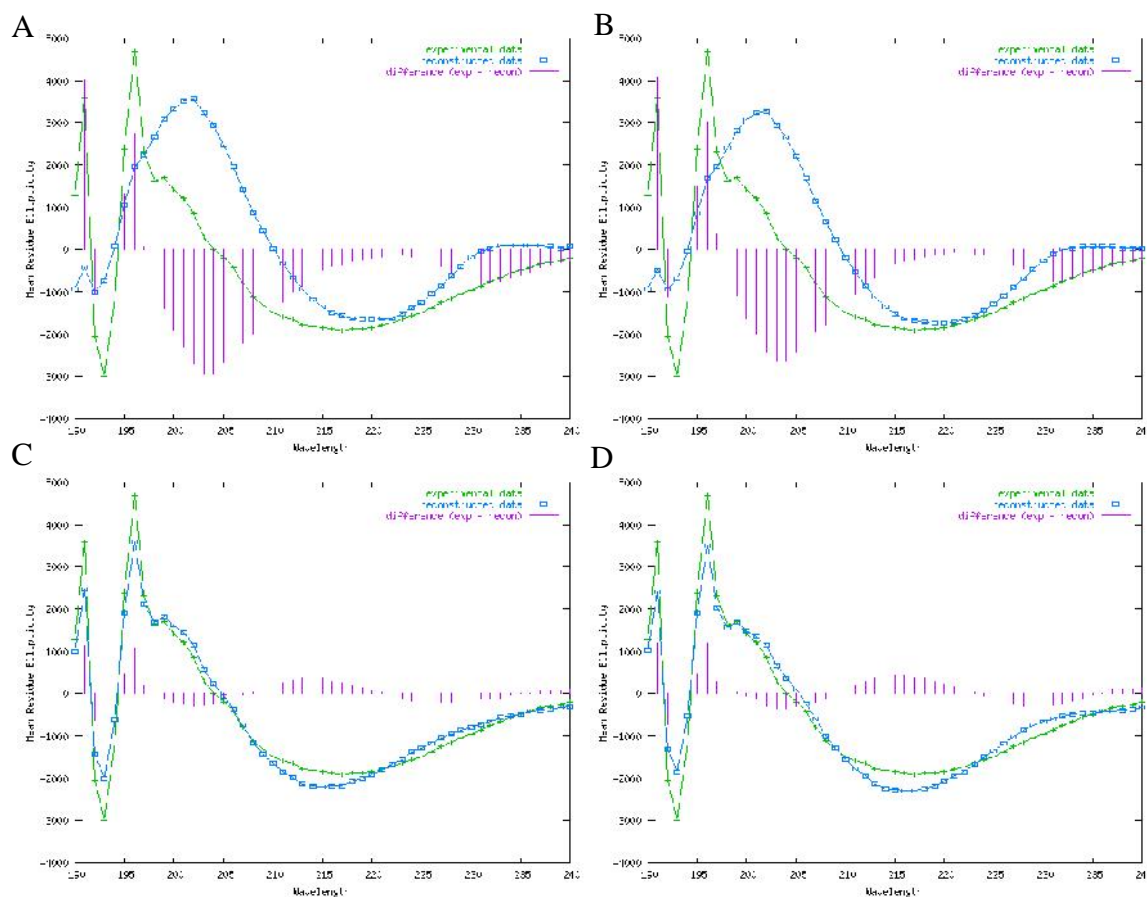
**Figure 5.4. CD spectra of *E. coli* OmpA TM domain.** White squares denote the N-terminal  $\beta$ -barrel domain from *E. coli* (Sugawara *et al.*, 1996). The negative peak occurs at approximately 215 nm and the positive peak at 200 nm. The peaks occur at wavelengths expected for predominantly  $\beta$ -stranded proteins (Greenfield, 2006). The white and black circles represent the CD spectra of whole OmpA purified in octyl-POE and SDS respectively.

CD data from BAPKO\_0422 was collected using a JASCO J-810 spectropolarimeter at the University of York as described in section 2.2.14 at two different concentrations (Figure 5.5 & Figure 5.7) in an effort to obtain as much data as possible. Data was processed using the online web server DichroWeb (Lobley *et al.*, 2002, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008) which uses a selection of algorithms to calculate the % secondary structure present. The algorithms CDSSTR and SELCON3 were chosen for analysis. CDSSTR discounts proteins from the reference sets that are not relevant to the experimental protein (Compton and Johnson, 1986), while SELCON3 analyses and refines the results as analysis progresses (Sreerama *et al.*, 1999). Both of the algorithms used a set of reference spectra for comparative analysis. There were several reference sets to choose from, sets 4 and 7 were selected as they were optimised for data in the range 190 - 260 nm (Janes, 2008), the same range of data which was acquired for BAPKO\_0422.



**Figure 5.5. CD of 0.33 mg/ml BAPKO\_0422.** Spectra collected at 20 °C in 0.1 % (w/v) LDAO, 0.3 M NaCl, 30 mM Tris-HCl, pH 8. Once the H[T] voltage crosses 600 V (dashed green line) the data at lower wavelengths becomes unusable as the data greatly varies. Data was acquired at a protein concentration of 0.33 mg/ml in 0.1 % (w/v) LDAO, 0.3 M NaCl, 30 mM Tris-HCl, pH 8) between wavelengths of 190 - 260 nm, however data is only useable above 198 nm. The overall shape of the experimental data up to this point (195 – 260 nm) is characteristic of  $\alpha$ -stranded protein and very similar to that of *E. coli* OmpA (Figure 5.4). Data was acquired in triplicate, shown is a representative trace.

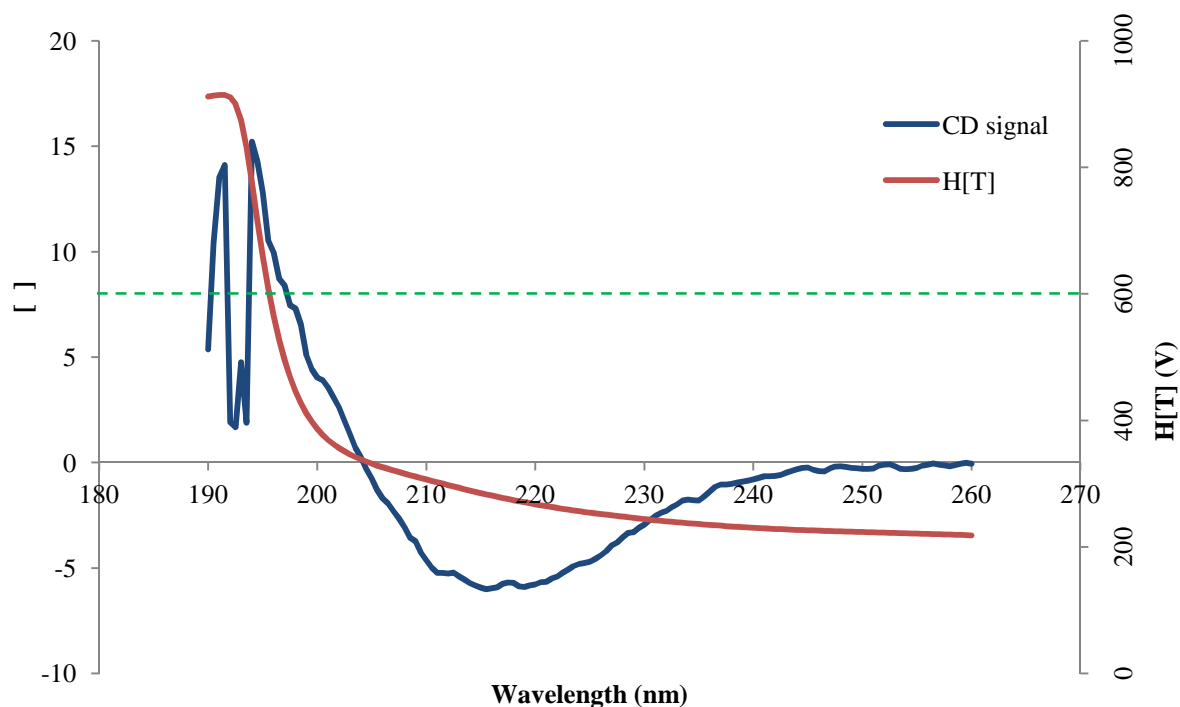
The spectra produced (Figure 5.5) has the same shape and has peaks at similar wavelengths as the *E. coli* N-terminal domain (Figure 5.4). The experimental data was submitted to the online server Dichroweb and analysed using algorithms SELCON3 and CDSSTR with reference datasets 4 and 7. The results are displayed below (Table 5.1), with the predicted fit of the calculated secondary structure percentages plotted with the experimental data (Figure 5.6).



**Figure 5.6. Plots of BAPKO\_0422 DichroWeb analysis.** Green lines show experimental data, blue lines show data calculated by the algorithms to fit the calculated secondary structure. The pink lines are the difference between the two spectra. **(A)** and **(B)** are plots from SELCON3 analysis using reference datasets 4 and 7 respectively. Each reference set is comprised of a majority of soluble globular proteins, with few transmembrane  $\alpha$ -barrel proteins. The experimental data varies greatly from the calculated spectra, suggesting a poor fit of the data (Whitmore and Wallace, 2004). **(C)** and **(D)** are plots produced by CDSSTR using reference datasets 4 and 7 respectively (Janes, 2008). Whilst there is a difference between the experimental and calculated data, the difference is not as large as that between the experimental and expected data in **(A)** and **(B)**, which suggests a more accurate calculation of % secondary structure.

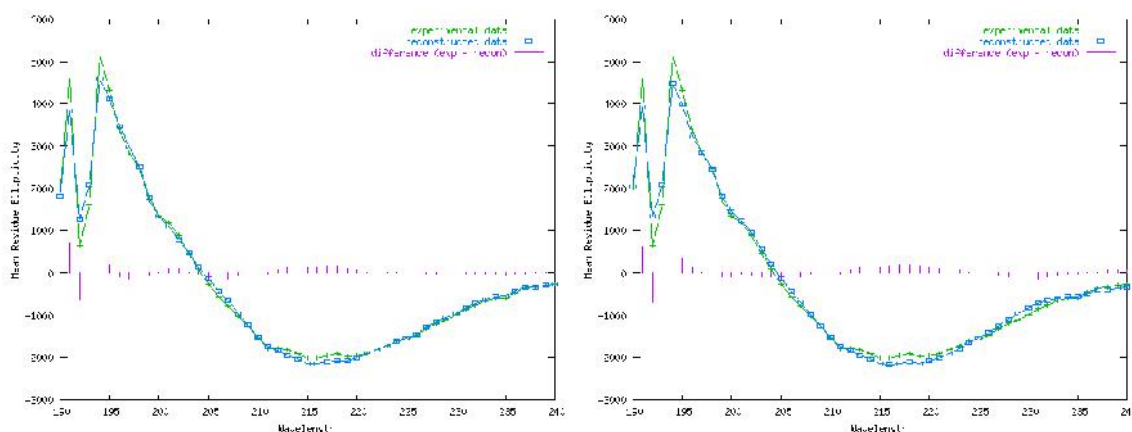
Each algorithm produced a theoretical CD spectra for a protein with the % secondary structure calculated (Figure 5.6). CDSSTR was preferred as this discounted reference sets dissimilar to OmpA;  $\alpha$ -barrel proteins are under-represented in the reference sets, hence CDSSTR analysis should be more accurate. SELCON3 was used as a comparison which took into account all reference datasets.

The H[T] exceeded 600 V before data at 190 nm had been obtained and so data was acquired at a second concentration; BAPKO\_0422 was diluted to 0.166 mg/ml in an effort to acquire data at lower wavelengths and run again (Figure 5.7).



**Figure 5.7. CD spectra of 0.166 mg/ml BAPKO\_0422.** Spectra acquired at 20 °C in 0.1 % (w/v) LDAO, 0.3 M NaCl, 30 mM Tris-HCl, pH 8. The dashed green line at 600 V highlights the point at which the CD signal becomes unreliable. The CD data obtained up to that point is indicative of  $\beta$ -stranded secondary structure with the peaks at approximately the same wavelengths as the N-terminal domain from *E. coli* OmpA.

As with the spectra obtained at 0.33 mg/ml, (Figure 5.5), this spectra has the same shape and has peaks at similar wavelengths as the *E. coli* N-terminal domain (Figure 5.4). The data obtained were analysed using the online server DichroWeb using the algorithm CDSSTR in conjunction with reference sets 4 & 7 (displayed along with analysis of data acquired at 0.166 mg/ml in Table 5.1). Analysis of the data plots also produced comparisons of the calculated spectra to the experimental spectra (Figure 5.8).



**Figure 5.8.** CDSSTR analysis of BAPKO\_0422 at 0.166 mg/ml. The experimental data is a very close fit to the Mean residue ellipticity calculated from the data, suggesting a good fit of the results.

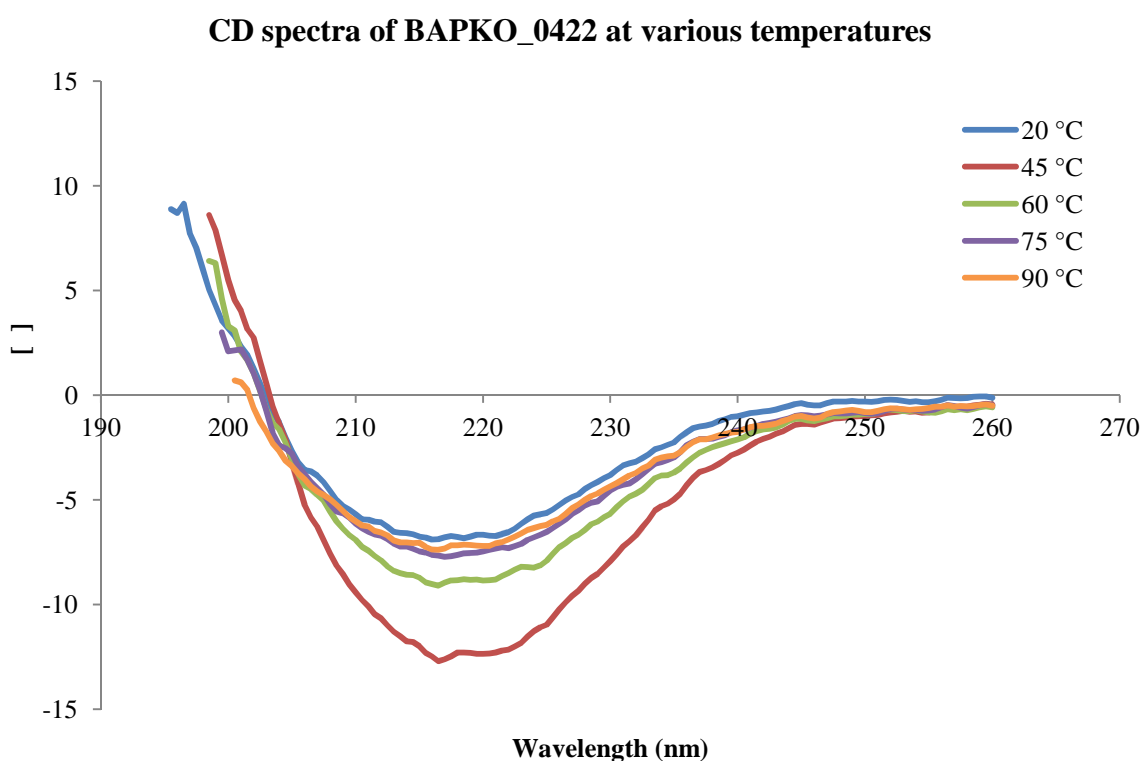
The results from 0.33 mg/ml and 0.166 mg/ml data were averaged (Table 5.1).

Conc (mg/ml)	Algorithm	Reference set	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
0.33	SELCON	4	-0.012	0.012	0.297	0.189	0.226	0.244	0.956
	SELCON	7	-0.010	0.013	0.295	0.190	0.229	0.239	0.956
	CDSSTR	4	0.00	0.05	0.31	0.14	0.20	0.30	1
	CDSSTR	7	0.00	0.02	0.26	0.14	0.21	0.36	0.99
	Average	-	-0.006	0.024	0.291	0.165	0.216	0.286	0.976
	<b>Total CDSSTR 0.33</b>		<b>0.035</b>		<b>0.425</b>		<b>0.205</b>	<b>0.33</b>	<b>-</b>
0.166	CDSSTR	4	0.01	0.07	0.23	0.14	0.22	0.33	1
	CDSSTR	7	0.00	0.03	0.23	0.13	0.18	0.41	0.98
	Average	-	0.005	0.05	0.23	0.135	0.20	0.37	0.99
	<b>Total 0.166</b>		<b>0.055</b>		<b>0.365</b>		<b>0.20</b>	<b>0.37</b>	<b>-</b>
-	<b>Averaged CDSSTR data</b>		<b>0.045</b>		<b>0.395</b>		<b>0.203</b>	<b>0.35</b>	<b>0.99</b>

**Table 5.1. Results from DichroWeb analysis.** The helix and strand columns are split into regular (1) and distorted (2) categories (Sreerama *et al.*, 1999). Despite SELCON3 being expected to give different results to CDSSTR, the percentages are in reasonable agreement with very low helical content and a large fraction of  $\beta$ -strand/turn. Nevertheless SELCON3 results have not been included in the average data. Averaging the 0.33 mg/ml data calculated by CDSSTR reveals that over 40 % of the structure is  $\beta$ -stranded and approximately a third is unordered. Analysis of 0.166 mg/ml data using CDSSTR and reference sets 4 and 7 calculated a smaller percentage of  $\beta$ -strand and higher percentage of turns and unordered secondary structure compared to data at 0.33 mg/ml.

Analysis of the calculated secondary structure percentages found that BAPKO\_0422 is approximately 40 %  $\alpha$ -sheet, 20 % turn and 35 % unordered. This is consistent with an OM-spanning protein.

OM-spanning  $\alpha$ -barrel proteins are expected to remain stable above 100 °C (Sugawara *et al.*, 1996), hence CD spectra were obtained at increasing temperatures (45 °C to 90 °C in 15 °C increments) to observe the thermal stability of the protein (Figure 5.9 & Table 5.2).



**Figure 5.9. CD spectra of BAPKO\_0422 at various temperatures.** Protein analysed at a concentration of 0.166 mg/ml and at five different temperatures. The data sets are cut off where the H[T] exceeded 600 V.

Each of the spectra has a shape similar to that of the N-terminal OmpA domain (Figure 5.4) with peaks at the same wavelengths.

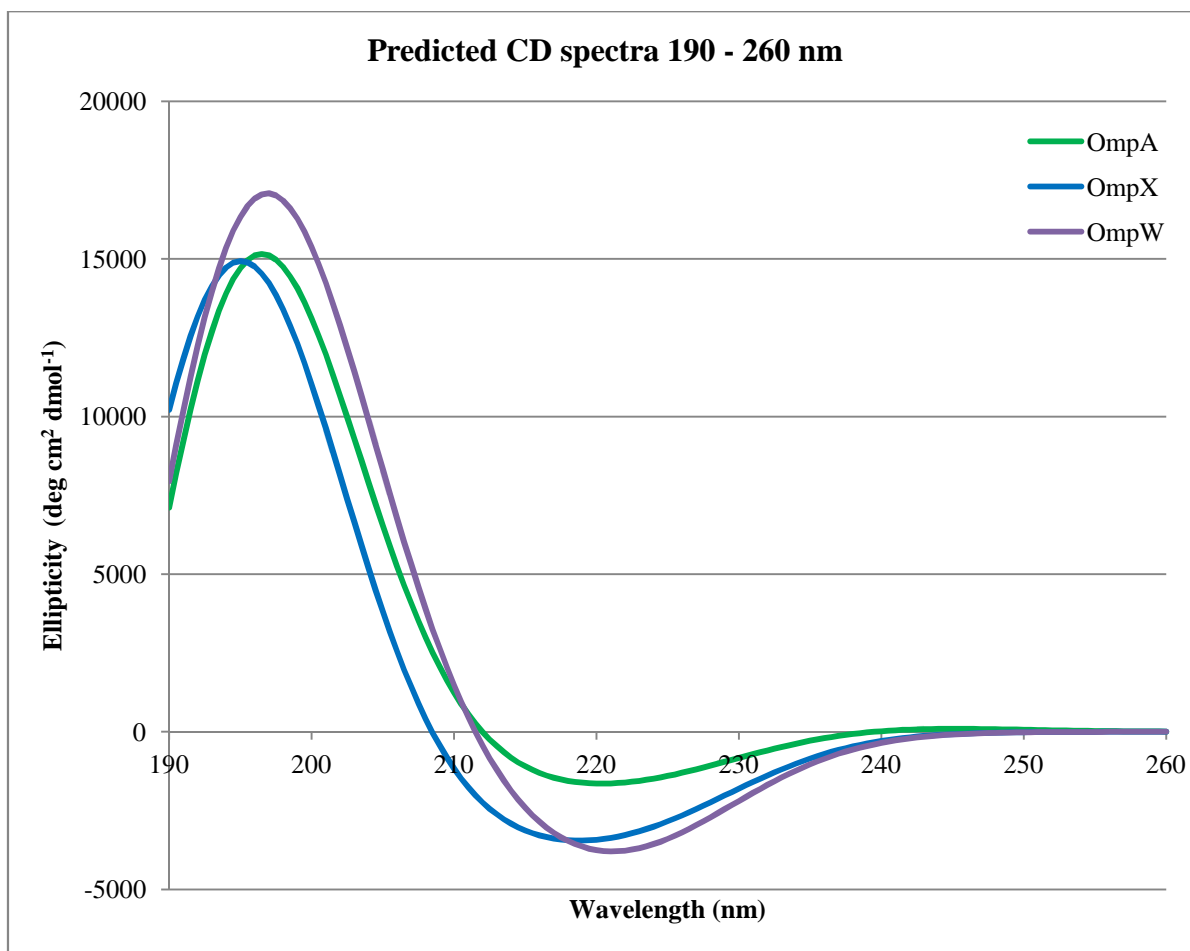
Temp. (°C)	Reference set	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
20	4	0.00	0.05	0.31	0.14	0.20	0.30	1
20	7	0.00	0.02	0.26	0.14	0.21	0.36	0.99
45	4	0.01	0.05	0.26	0.13	0.22	0.32	0.99
45	7	0.00	0.02	0.25	0.13	0.20	0.39	0.99
60	4	0.00	0.06	0.25	0.13	0.25	0.30	0.99
60	7	0.00	0.02	0.20	0.11	0.18	0.47	0.98
75	4	0.01	0.07	0.24	0.12	0.27	0.28	0.99
75	7	0.00	0.03	0.21	0.11	0.21	0.42	0.98

**Table 5.2. DichroWeb analysis of BAPKO\_0422 at varying temperatures.** The data obtained at each temperature was analysed using CDSSTR using reference sets 4 and 7. The 20 °C data is that analysed previously (Table 5.1). Data obtained at 90 °C could not be analysed by the online server. The percentage  $\alpha$ -strand has decreased by approximately 13 % once the protein reaches 75 °C.

Overall there is very little change in secondary structure with heating; only a gradual decrease in the %  $\alpha$ -stranded structure with heating. This is not unexpected as the purified N-terminus of *E. coli* OmpA has been shown to be stable at 100 °C (Sugawara *et al.*, 1996).

The reference datasets 4 and 7 used by SELCON and CDSSTR are based predominantly on soluble globular proteins, with few representative membrane proteins. As BAPKO\_0422 is predicted to be a membrane spanning beta barrel, it would be informative to compare the CD spectra with a similar protein. To this end theoretical CD spectra of *E. coli* OmpA, OmpW and OmpX were calculated using the DichroCalc webserver (Bulheller and Hirst, 2009). The wavelength range was set to 190 – 260 nm with 0.5 nm steps to match the experimental data obtained from BAPKO\_0422 (Figure 5.10).





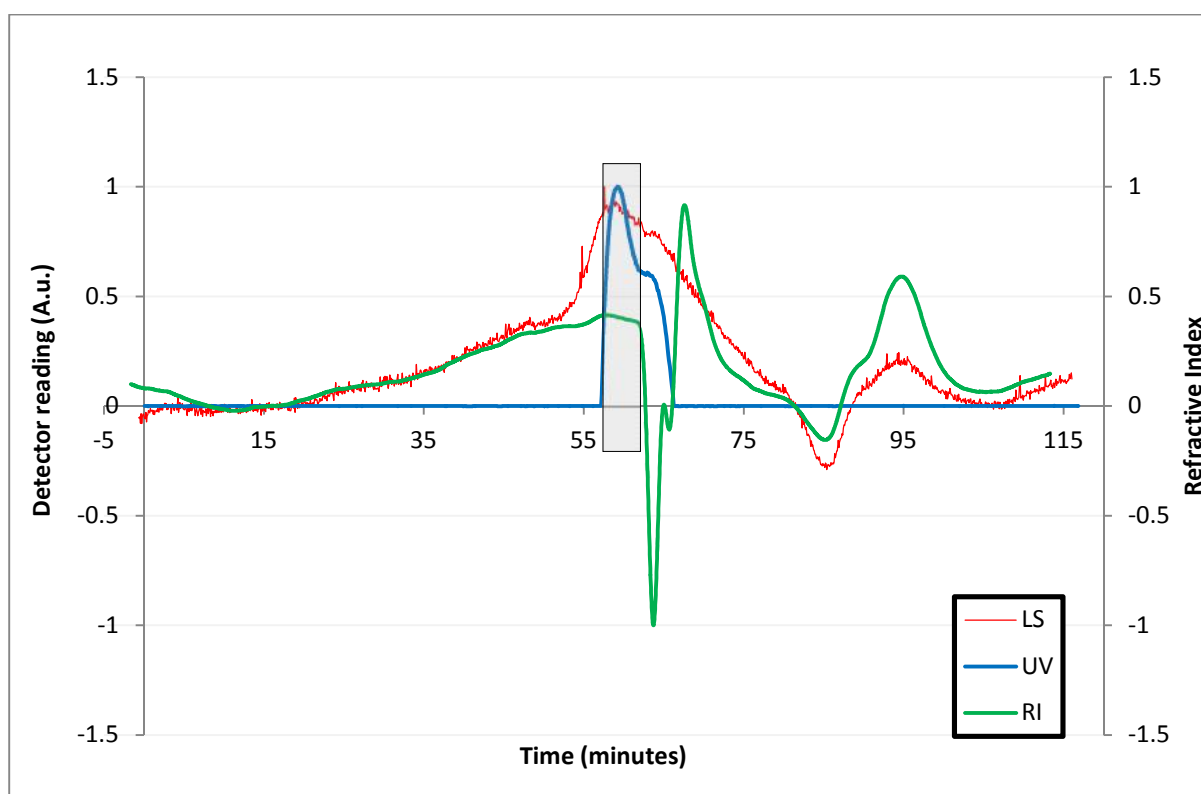
**Figure 5.10. Predicted CD spectra of known OM proteins from *E. coli*.** Spectra generated from the .pdb files for 1BXW (OmpA), 1QJ8 (OmpX) and 2F1T (OmpW) using the online webserver DichroCalc (Bulheller and Hirst, 2009). Compared to the experimental BAPKO\_0422 data collected at 0.166 mg/ml (Figure 5.7) the overall shape of the curve is very similar, with the most negative value approximately one third the value of the peak.

The predicted spectra all have positive and negative peaks at approximately the same positions as the N-terminal *E. coli* domain (Figure 5.4) and BAPKO\_0422 (Figure 5.5, Figure 5.7 and Figure 5.9); OmpX however is predicted to have peaks (both positive and negative) a couple of nm lower than both OmpA and OmpW.

## 5.2 SEC-MALLS

Size Exclusion Chromatography-Multi Angle Laser Light Scattering (SEC-MALLS) was used to determine the oligomeric state of BAPKO\_0422 in solution. This is achieved using the protein concentration ( $c$ ), a constant ( $K$ ), the difference between intensity of light at zero angle and scattered ( $LS$ ) and  $dn/dc$  (the refractive index increment of the protein which relates changes in refractive index ( $RI$ ) to protein concentration).  $K$  is dependent on the  $RI$  of blank buffer, wavelength used, the scattering angle and distance between scattering and the detector. The  $dn/dc$  value for BAPKO\_0422 in detergent was not known, hence the default value for soluble proteins of 0.187 ml/g was used.

A sample of BAPKO\_0422 was concentrated to 4 mg/ml in 0.1 % (w/v) LDAO, 0.3 M NaCl, 50 mM Tris-HCl, pH 8.0 and centrifuged at 14,500 rpm in a bench top centrifuge for 15 minutes to remove any protein aggregates. 150  $\mu$ l of sample was loaded onto the SEC-MALLS system.



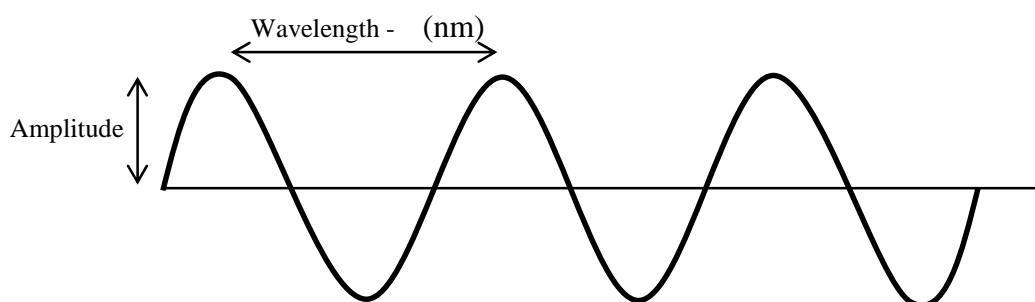
**Figure 5.11. SEC-MALLS trace of BAPKO\_0422.** Data was acquired at a concentration of 4 mg/ml in SEC buffer. The gray highlighted area was selected for analysis. Light scattering is measured on the primary axis along with UV absorbance at 280 nm.

Separation was achieved using three Agilent PL Aquagel-OH (40, 50 and 60) SEC columns linked sequentially. The UV trace showed elution at approximately 58 minutes (Figure 5.11).

The span of the peak from 58.006 to 60.256 minutes was analysed assuming a  $dn/dc$  value of 0.187 ml/g with manually adjusted background levels for all light scattering detectors. This produced a molecular weight of 22.96 kDA ( $\pm 9\%$ ), which corresponds with the expected molecular weight of the protein including His-tag as calculated using the ExPASy ProtParam tool - 22.68 kDA (Gasteiger *et al.*, 2005). The polydispersity was calculated to be 1.001 ( $\pm 12\%$ ), showing that the protein is monomeric under the conditions tested.

### 5.3 Small Angle X-Ray Scattering

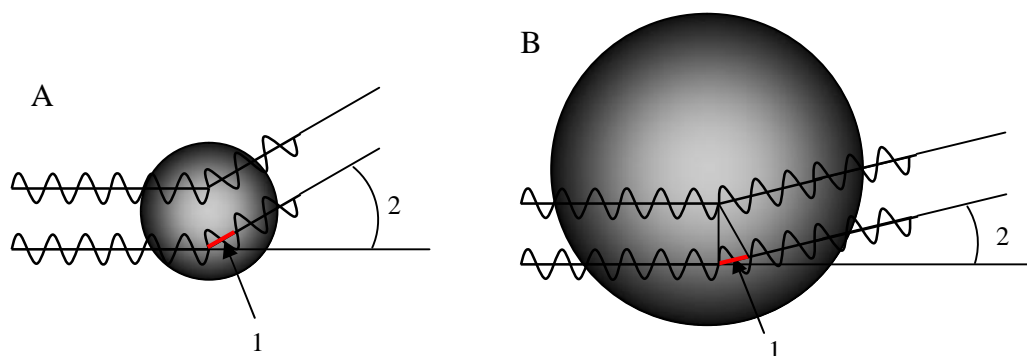
X-Rays are plane waves, which have a regular waveform with a distinctive *sinusoidal* shape (Figure 5.12) (Blow, 2002). When X-Rays pass through matter, scattering by electrons occurs. The electrons resonate with the frequency of X-rays passing through the objects and emit coherent secondary photons or wavelets (Jacques and Trewhella, 2009) which interfere with each other (Glatter and Kratky, 1982). The electron may oscillate with the same frequency as the X-ray and radiate an X-Ray photon of the same energy and wavelength in a random direction, whilst remaining in phase. The two possible outcomes are constructive and deconstructive interference (or coherent and incoherent scattering) (Jacques and Trewhella, 2009). Interaction of an X-ray with electrons may also result in atomic transitions within the atom and the release of photons with lower energy than the original X-ray, leading to incoherent scattering (Blow, 2002).



**Figure 5.12. The sine wave of an X-ray.** A basic representation of an uninterrupted X-ray.

Both X-Ray crystallography and SAXS take advantage of an inverse relationship between scattering and the molecules. In structure determination by X-Ray crystallography the inverse relationship is between the spacing of unit cells in the crystal (real lattice) and the spacing of reflections on the detector (reciprocal lattice); enabling calculation of the atomic distances in real space (Rhodes, 2000). In SAXS the particle size is inversely proportional to the angle at which X-Rays are scattered, facilitating calculation of the particle size (Glatter and Kratky, 1982).

Small Angle X-Ray Scattering (SAXS) is an analytical technique used to study structural features of colloidal size (Glatter and Kratky, 1982). As stated the scattering of X-rays by particles in solution or a solid object is the product of an inverse relationship between the size of the particle and the angle of scattering (Figure 5.13) (Glatter and Kratky, 1982). As SAXS is conducted in solution, the rotational averaging of X-Ray scattering limits the resolution of the data obtained, however analysis of SAXS data is capable of producing precise information about the size and shape (Jacques and Trehwella, 2009).



**Figure 5.13. Scattering angle relative to particle size.** **A)** A small spherical particle. **B)** A large spherical particle. Assuming the waves scattered from the two points at an angle of  $2$  have a distance of  $1$ . In a larger molecule, the distance of  $1$  will occur at narrower scattering angles, hence the inverse relationship between particle size and scattering angle. Image based on (Glatter and Kratky, 1982). In homodisperse solutions, the scattering intensities of the particles will accumulate.

The scattering pattern produced by SAXS is described by the intensity ( $I$ ) as a function of the amplitude of the scattering vector ( $q$ ) (Jacques and Trehwella, 2009).

$$q = \frac{4\pi \sin \theta}{\lambda}$$

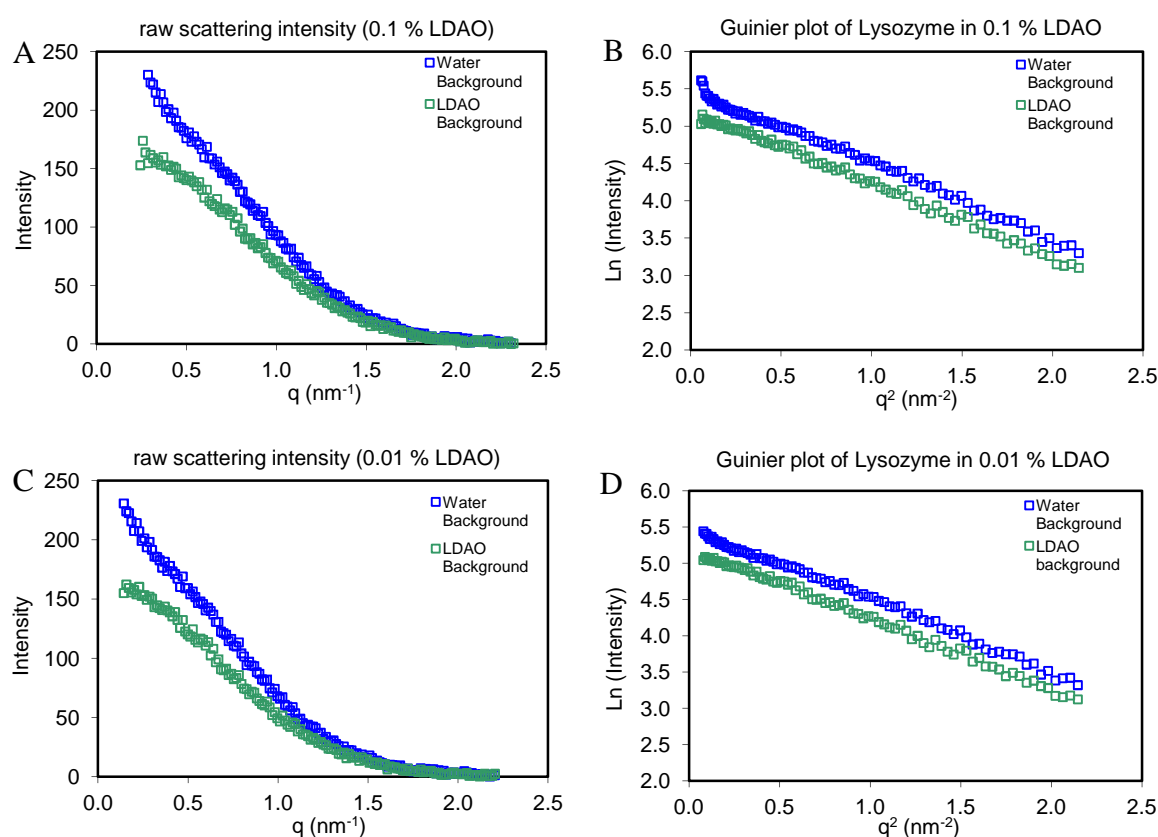
**Equation 4. Small angle scattering as a function of  $q$ .**  $4 \sin \theta / \lambda$  represents the intensity of small angle scattering ( $I$ ).

### 5.3.1 Control analysis (lysozyme)

A control experiment using 10 mg/ml lysozyme was used as a protein standard to establish that proper background subtraction has been performed. Poor background subtractions can lead to significant sources of error (Jacques and Trehwella, 2009), in addition different methods of background subtraction have different sources of error, hence multiple background subtractions should be used. The background was measured using different buffers (both with and without LDAO) and through the use of lysozyme as a control standard. Close agreement in the background scattering in conjunction with a linear Guinier plot and

“well-behaved  $P(r)$  function” is strongly indicative of a monodisperse system (Jacques and Trehella, 2009).

In order to determine the effect of detergent on the scattering pattern data was collected from buffer containing 0 % (w/v), 0.1 % (w/v) and 0.01 % (w/v) LDAO (buffer compositions Table 2.2). The salt background (0 % detergent) scattering pattern was subtracted from data sets of both buffers and so the contribution of salt to scattering patterns was observed. This was found to be negligible. Scattering data produced by 10 mg/ml Lysozyme in two buffers with different concentrations of LDAO (0.1 (w/v) & 0.01 % (w/v)) (Figure 5.14 A & B and C & D respectively) was obtained in duplicate over a period of 24,000 seconds.



**Figure 5.14. Scattering produced by 10 mg/ml Lysozyme in 0.1 and 0.01 % (w/v) LDAO.** All data is averaged after software background scattering subtraction. Blue data has had the scattering from a water background automatically subtracted, green data has had background scattering from buffer subtracted. **A)** 0.1 % (w/v) LDAO - Plot of raw scattering intensity of lysozyme. **B)** 0.1 % (w/v) LDAO - Guinier plot, **C)** 0.01 % (w/v) LDAO - Plot of  $q$  vs scattering intensity. **D)** 0.01 % (w/v) LDAO - Guinier plot data was acquired in duplicate over a period of 24,000 seconds per acquisition, data plotted on both graphs is the average of both acquisitions minus the background scattering recorded (not shown). Plots were normalised through the subtraction of the average scattering intensities at high  $q$ . All data sets shown were collected in duplicate, averaged and corrected for thermal background. There is a visible difference at low  $q$  in the intensity of scattering produced by lysozyme when different background data sets are used, highlighting that the detergent contributes to observed scattering and must be accounted for. Analysis using water analysis in **B** appears to have a small upturn at low  $q^2$ , suggesting aggregation of the sample; this may be an artefact produced by not accounting for detergent scattering.

### 5.3.1.1 Calculating the radius of gyration (R<sub>g</sub>) of lysozyme

From the scattering pattern, many analyses can be conducted; primarily calculation of the radius of gyration (R<sub>g</sub>). The R<sub>g</sub> provides information about the mass distribution within a particle and is defined as the root-mean-squared distance of all elemental scattering volumes from their centre of mass weighted by scattering densities (Jacques and Trewhella, 2009); providing a measure for the spatial size of the particle (Glatter and Kratky, 1982). Simply put, the R<sub>g</sub> is the distance in a rotating system between the point at which it is rotating and the point at which the mass about the axis will be concentrated. Objects with the same volume but different shapes will have different R<sub>g</sub> values (Jacques and Trewhella, 2009). The R<sub>g</sub> of a spherical molecule can be calculated using the slope of the Guinier plots by application of the Guinier approximation at low *q* using the following equation:

$$R_g = \sqrt{-3m}$$

**Equation 5. Calculating the radius of gyration for a spherical object from a Guinier plot.** *m* = the slope of the line. The radius of gyration for a sphere was calculated as the protein-detergent complex was calculated as this was deemed to give the most accurate result. As the scattering patterns are rotationally averaged the best estimate for radius of gyration would be one that affords equal weighting in all directions (spherical R<sub>g</sub>). This gives the R<sub>g</sub> in nm so a 10 x multiplication is required to convert to Å

From the raw data, information about the state of the solution is available; any upturn in the data at low *q*<sup>2</sup> on a Guinier plot is a good early indicator of protein aggregation, whilst a downturn can be symptomatic of interparticle interference (Jacques and Trewhella, 2009). The R<sub>g</sub> for lysozyme was calculated using the slope of the Guinier plots (Table 5.3) and Equation 5. Calculation of the R<sub>g</sub> from lysozyme was found to vary depending on the background scattering subtracted (Table 5.3).

Buffer	Background scattering	R <sub>g</sub> (Å)
0.1 % (w/v) LDAO	Water	19.0
0.1 % (w/v) LDAO	Detergent	15.6
0.01 % (w/v) LDAO	Water	17.4
0.01 % (w/v) LDAO	Detergent	15.6

**Table 5.3. Calculated R<sub>g</sub> of lysozyme.** The radius of gyration for lysozyme was calculated using the slope of the Guinier plots (Figure 5.14 B & D) after any upturn in data and Equation 5. The R<sub>g</sub> was found to be the same when using each detergent background-subtracted data. However when the R<sub>g</sub> is calculated using water background-subtracted data the R<sub>g</sub> is different depending on the concentration of LDAO. This highlights the necessity to compensate for the scattering produced by detergent molecules/micelles.

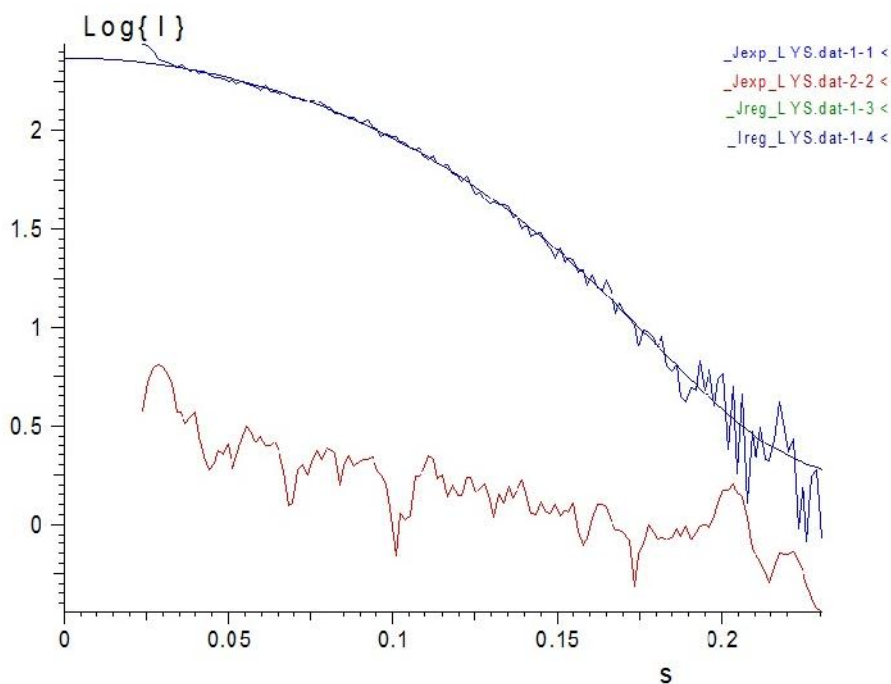
The use of detergent-background subtracted data gives a more precise value, remaining the same despite differing buffer concentrations, confirming the detergent contribution to the scattering pattern. The  $R_g$  for lysozyme calculated by PRIMUS (part of the ATSAS data package) in 0.01 % (w/v) LDAO detergent background-subtracted was 15.4 Å (data not shown), and 15.5 Å for lysozyme in 0.1 % (w/v) LDAO (data not shown) both of which are in close agreement with the detergent background subtracted data in Table 5.3 above. Averaging the four  $R_g$  values with detergent background subtraction gives an  $R_g$  for lysozyme in LDAO of  $15.53 \pm 0.02$  Å. The error quoted is the standard error of the mean. The  $R_g$  for lysozyme obtained from both detergent background-subtracted samples is in relatively close agreement with previous experimental data; 16.3 Å at 20 °C (Arai and Hirai, 1999).

### **5.3.1.2 Evaluation of the particle distance distribution function**

The particle distribution function -  $P(r)$  - is an indicator of the quality of the data and amount of protein aggregation. The  $P(r)$  is obtained through a Fourier transform of the raw scattering data (Jacques and Trewhella, 2009). Aggregated samples may give one or more small peaks at high  $R$  values and samples experiencing interparticle interference may reach 0 at  $D_{\max}$  (distribution at longest particle dimensions) values smaller than expected.  $D_{\max}$  is selected upon input of the data into GNOM (defined as  $R_{\max}$ ) and if  $D_{\max}$  reaches 0 abruptly it is likely that the value needs increasing. The  $P(r)$  function is also an indicator of interparticle distances (particle overlap); if there are multiple peaks or a shoulder on a peak this is an indicator that the particles are overlapping in solution (Roe, 2000).

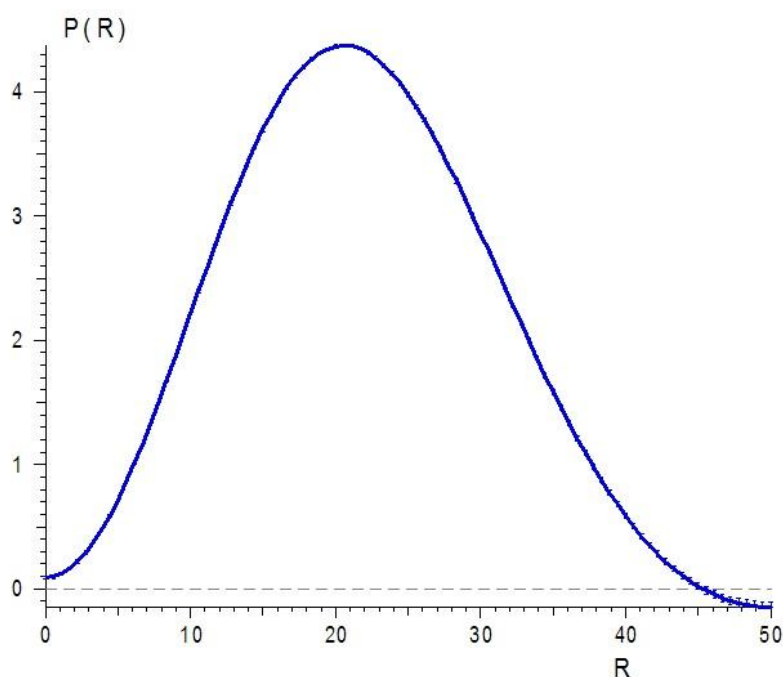
GNOM is an indirect transform program for interpretation of small angle scattering data (Svergun, 1992). The data submitted to PRIMUS was run through GNOM and plotted as a reciprocal space fit (Figure 5.15). Default parameters were used except those stated in section 2.2.15.2. The  $R_g$  calculated from the reciprocal space fit (Figure 5.15) is 16.39 Å, which is in very strong agreement with the published value of 16.3 Å (Arai and Hirai, 1999).



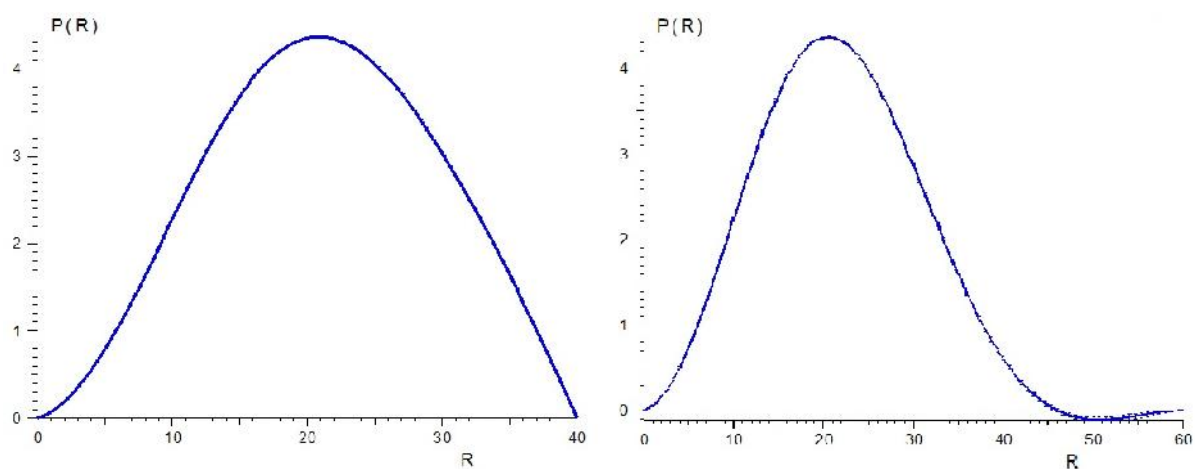


**Figure 5.15. GNOM reciprocal space fit of 10 mg/ml lysozyme.** Data collected in 0.1 % (w/v) LDAO, 0.3 M NaCl, 30 mM Tris-HCl, pH 8). The smooth blue line (Ireg\_LYS.dat-1-4) is the GNOM fit to the data, the other blue line is the experimental data (Jexp\_LYS.dat-1-1) and the red line is the experimental input points used for regularisation (Jexp\_LYS.dat-2-2) (Svergun, 1992). 'q' is designated 's' in the program. The fit of the GNOM plot to the experimental data is in close agreement.

Scattering curves are used to generate the particle distance distribution function  $P(r)$  and size distribution function  $D(R)$  for mono and poly-disperse systems respectively; this conversion is capable of compensating for data that has been subjected to 'smearing' by distortion from the instruments. As all samples in this research were expected to be monodisperse (confirmed by the calculated  $P(r)$  functions)  $P(r)$  was used.  $P(r)$  is obtained by Fourier inversion of the scattering data (Glatter and Kratky, 1982). The conversion requires several parameters, including two that are important in obtaining a good  $P(r)$ : ALPHA and  $D_{\max}$ . GNOM automatically calculates various values for ALPHA (a regularisation parameter) and plots the distribution of various alpha values against the quality of the solution (Svergun, 1992). The larger the alpha, the more attention paid to the smoothness of the solution, and less attention to fitting the experimental data.  $D_{\max}$  as previously mentioned is the maximum expected particle size (or the distance between particles at which no overlap is expected (Roe, 2000)) Figure 5.16 & Figure 5.17 show the calculated  $P(r)$  for lysozyme using three different values for  $D_{\max}$  to determine the correct one for use.



**Figure 5.16.**  $P(r)$  of lysozyme in 0.1 % (w/v) LDAO with a  $D_{\max}$  of 50. Error bars have been automatically calculated and incorporated into the plot. There is no secondary peak at high  $R$  suggesting that there is either no or little aggregation in the sample. The plot reaches 0 at a steady rate, indicating that  $D_{\max}$  of 50 is a good estimate and the lack of further peaks suggests no particle overlap.

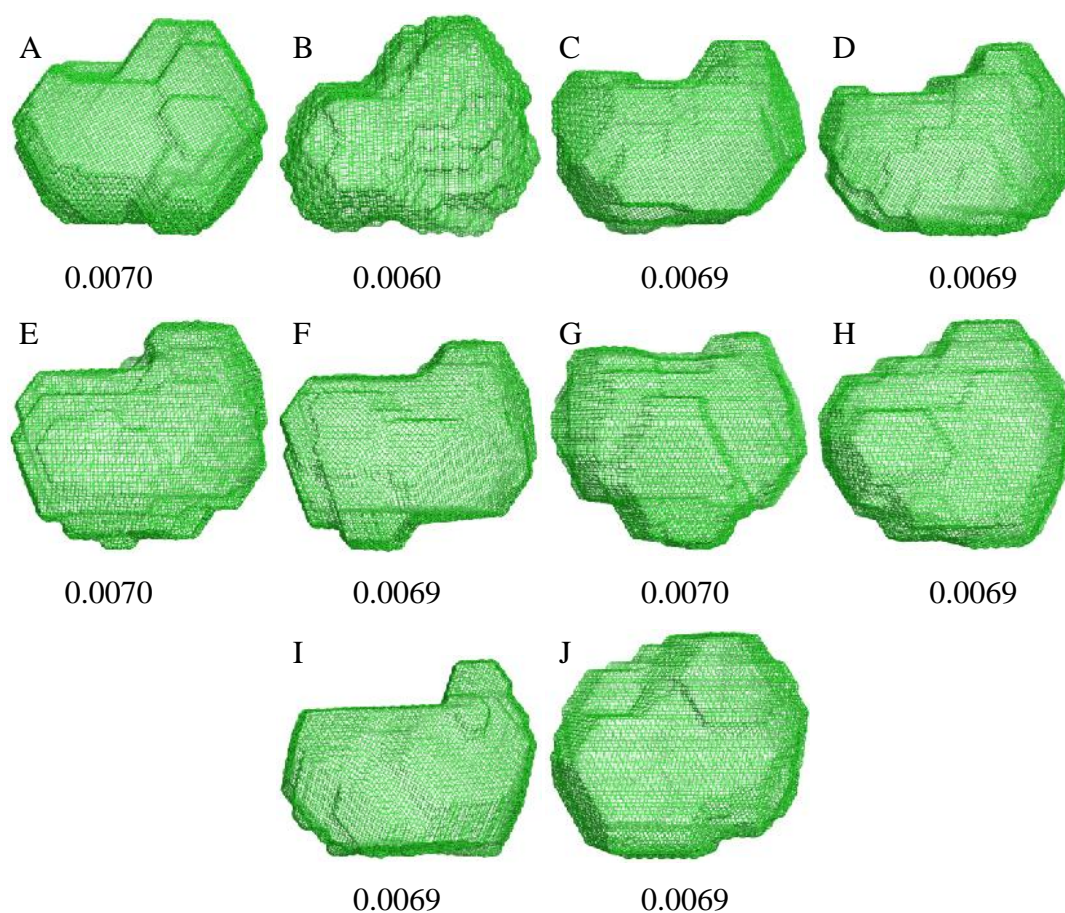


**Figure 5.17.**  $P(r)$  plots of lysozyme in 0.1 % (w/v) LDAO with  $D_{\max}$  values 40 & 60. A)  $D_{\max}$  of 40; B)  $D_{\max}$  of 60. With a  $D_{\max}$  value of 40 the plot reaches 0 abruptly, indicating that the value is too small; a value of 60 results in a curve that results in a negative  $P(r)$  at high values, indicating that this value is too large.

One of the final outputs is a rating for the solution calculated with regards to the data. Here the total estimate value calculated was 0.898, indicating a “GOOD” solution. There is a disagreement of 0.9 Å between the  $R_g$  calculated manually and by PRIMUS and that calculated by GNOM. As is visible in Figure 5.15 the data set has some values that do not fit the curve at low  $q$  (s), and the data in general becomes more varied as  $q$  approaches 0.2.

### 5.3.1.3 Generating the molecular envelope

The output file generated by GNOM from 10 mg/ml Lysozyme in SAXS 0.1 % (w/v) LDAO buffer was submitted to DAMMIF ten separate times to generate ten solutions for the molecular envelope. The resulting molecular envelopes calculated follow (Figure 5.18). Included is the value of the fit of the solution to the data; the closer to 0 the better the fit. A high value such as that obtained at the beginning of modelling ranges between 1.0 and 1.8.



**Figure 5.18. Molecular envelopes of lysozyme generated from 10 mg/ml lysozyme in SAXS 0.1 % (w/v) LDAO buffer.** Below each model is the fit of the solution. Each model has a very good fit value and all of the models are of similar shape with a region of density protruding at the top right of the molecular envelop as is viewed.

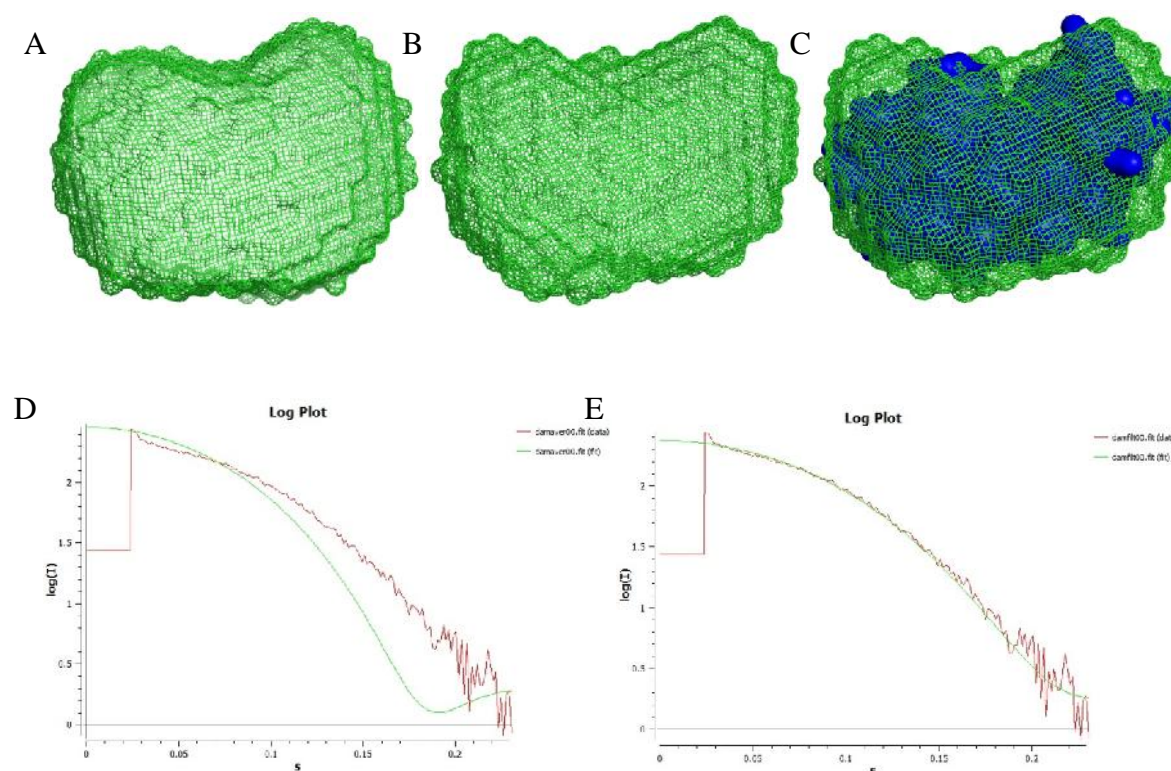
### 5.3.1.4 Evaluation of the model

DAMSEL and DAMSUP were used to evaluate the fit of molecular envelopes by comparison to an averaged solution created using all ten of the generated models (Table 5.4). If any models were too different from the average they were discarded and not used for the rest of the analysis.

Solution	Average DAMSEL comparison NSD	Include/discard for DAMSUP	DAMSUP superimposition NSD value
F	0.498	Reference	Reference
E	0.509	Include	0.498
I	0.510	Include	0.495
H	0.510	Include	0.494
D	0.514	Include	0.493
J	0.514	Include	0.488
B	0.516	Include	0.492
C	0.520	Include	0.495
A	0.521	Include	0.503
G	0.544	Discard	-

**Table 5.4. DAMSEL comparison of lysozyme molecular envelope solutions.** The overall mean NSD (section 2.2.14.2) for all models was 0.516 and the variance 0.012; giving NSD limits for accepting models of 0.492 and 0.540. All of the models with the exception of solution G fit within this range. All solutions except for G were included for further analysis. Each of the solutions shown in Figure 5.18 that were accepted by DAMSEL (include/discard column) were superimposed onto solution F (being used as a reference sample) and the NSD values relative to this reference were calculated.

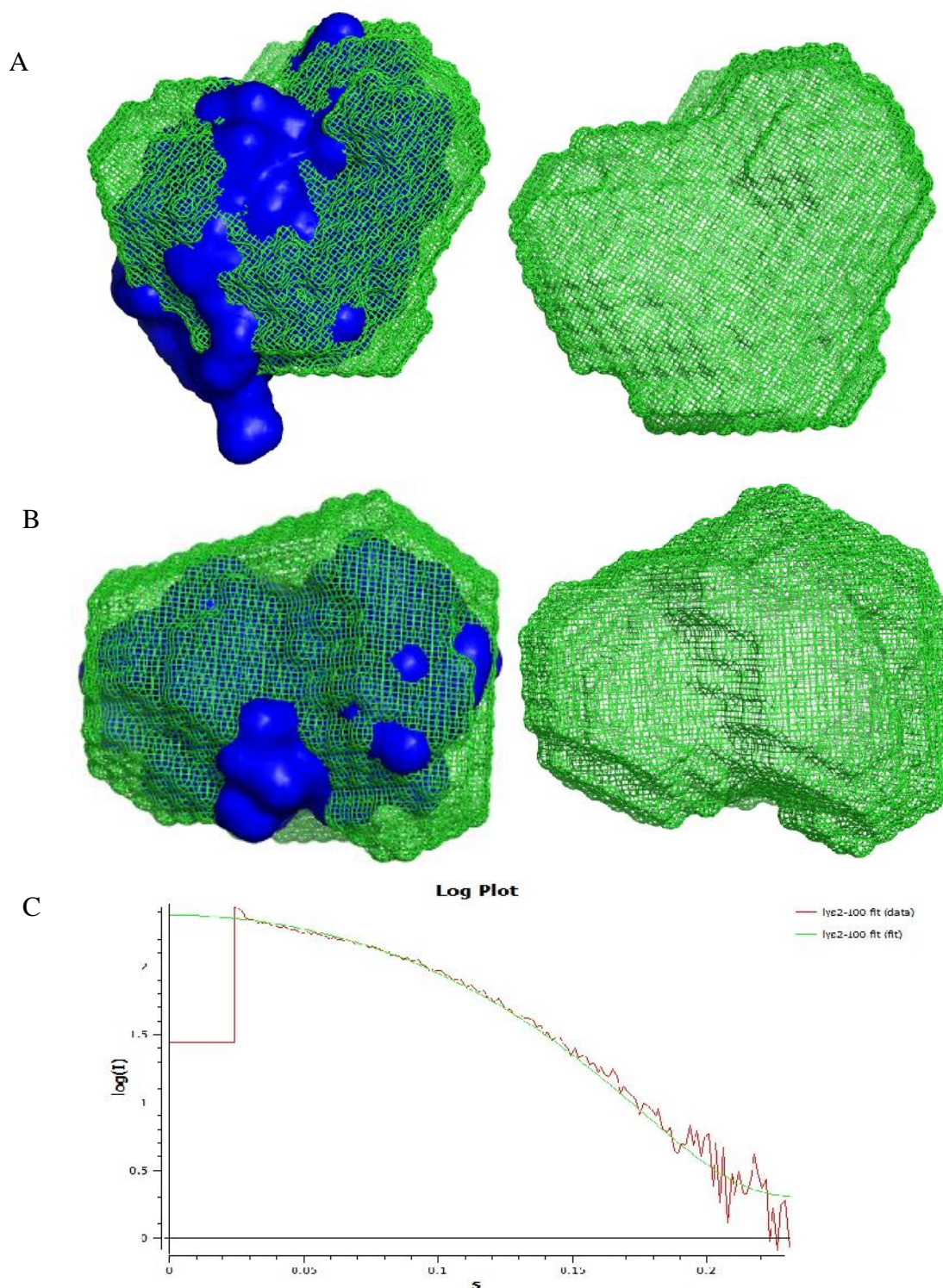
The DAMSUP file detailing the models to discard/include was used to produce an averaged molecular envelope (DAMAVER, Figure 5.19 A) before being refined to remove low occupancy regions (DAMFILT, Figure 5.19 B and C, C has the crystal structure of lysozyme added for comparison). The fit of the each solution was compared to the experimental data using CRY SOL (Figure 5.19 D and E).



**Figure 5.19. Lysozyme molecular envelope after refinement.** **A)** Averaged molecular envelop generated by DAMAVER, from the plot of the experimental data compared to the expected scattering from the DAMAVER solution. **D)** CRY SOL plot which shows the fit of the predicted scattering (green line) of molecular envelope (**A**) to the experimental data (red line). The varied difference in curve shows this is not an accurate representation of the data. **B)** and **C)** are DAMFILT refined molecular envelopes with and without the crystal structure of lysozyme. The molecule fits within the molecular envelope well with minimum empty space. **E)** CRY SOL plot showing the fit of the scattering (green line) of solution (**B**) to the data (red line). The calculated molecular envelope is predicted to produce a scattering pattern very similar to that obtained experimentally, indicating a good solution.

The DAMAVER model (Figure 5.19 A) was run through DAMSTART, which removed low occupancy and weakly interacting regions to prepare a .pdb file with a fixed core, which in turn was used by DAMMIN to generate the final molecular envelope (Figure 5.20 A & B, with and without the crystal structure).



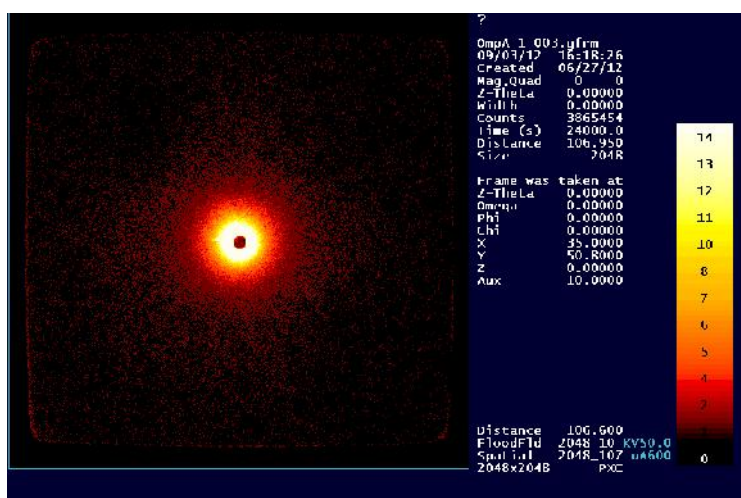


**Figure 5.20. DAMMIN refinement of the molecular envelope.** **A)** The generated molecular envelope from lysozyme data, with & without the crystal structure (1LYZ) showing the solvent accessible surface. **B)** Top down view of (A). **C)** CRY SOL plot of experimental data (red) with predicted scattering of the final molecular envelope (green). The generated molecular envelope fits the generated molecular envelope very well visually, as does the theoretical scattering curve to the experimental data.

The control analysis using lysozyme was a success. The scattering caused by detergent was compensated for by the software, the  $R_g$  calculated was very close to that in the literature and an accurate molecular envelope was generated.

### 5.3.2 Experimental analysis, BAPKO\_0422

Data for BAPKO\_0422 was collected at two different concentrations to provide information about the quality of the sample. If the sample is of good quality (e.g. monodisperse and devoid of interparticle interference) the calculated  $R_g$  should remain consistent regardless of protein concentration; an increase in  $R_g$  with an increase in concentration is indicative of protein aggregation, whilst a decrease in  $R_g$  with an increase in concentration is a sign of interparticle interference (Jacques and Trewhella, 2009). Figure 5.21 shows an example of the scattering pattern obtained from BAPKO\_0422 in SAXS 0.1 % (w/v) LDAO buffer.



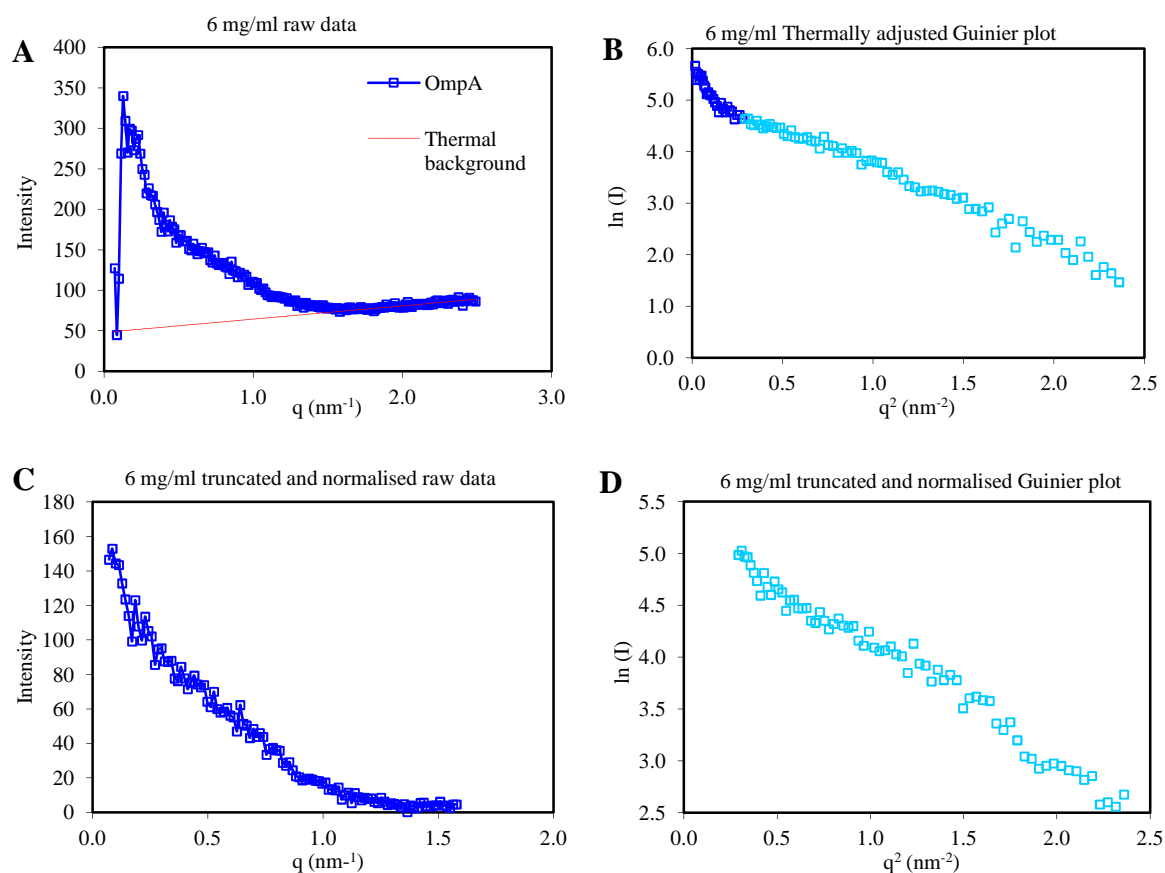
**Figure 5.21.** Scattering pattern produced by BAPKO\_0422 in solution. Data collected in house on a Bruker NANOSTAR over a period of 65,000 seconds at a concentration of 10 mg/ml in 0.1 % (w/v) LDAO SAXS buffer.

#### 5.3.2.1 Calculation of the $R_g$ for BAPKO\_0422

SAXS data was acquired for BAPKO\_0422 at concentrations of 6 and 10 mg/ml in 0.1 % (w/v) LDAO, 0.3 M NaCl, 30 mM Tris-HCl, pH 8.

Figure 5.22 shows raw data and Guinier plots from BAPKO\_0422 at a concentration of 6 mg/ml. The dataset is plotted twice; normalised using a thermal background (Figure 5.22 A) and normalised manually to 0 with the aggregation data upturn at low  $q$  and noise at high  $q$  removed (Figure 5.22 C). Each dataset has been converted to a Guinier plot (Figure 5.22 B and D respectively).

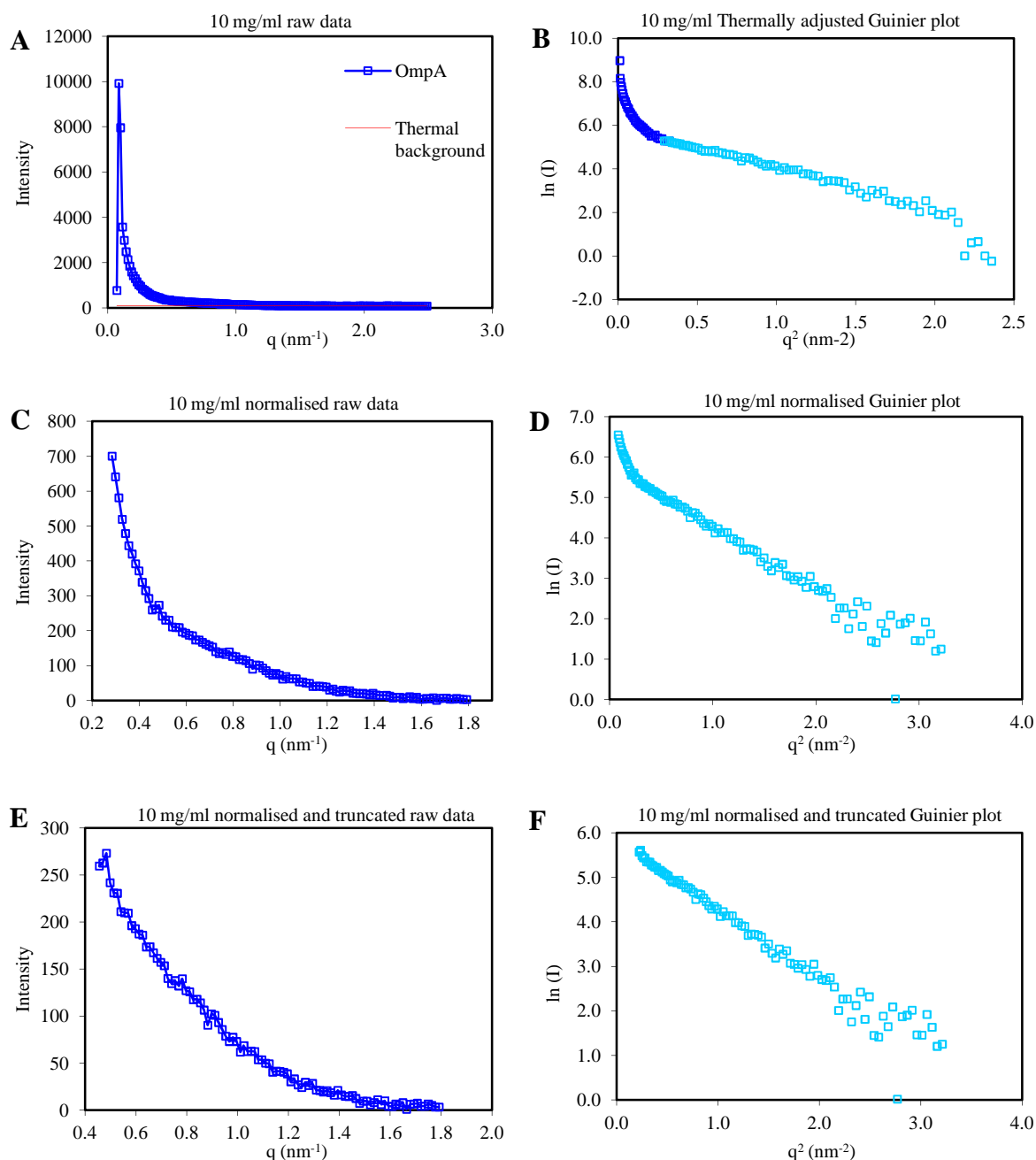
BAPKO\_0422 is predicted to be a barrel shaped molecule, however the  $R_g$  was calculated assuming a spherical molecule. However the  $R_g$  was calculated assuming the molecule was spherical to give an approximation of particle size.



**Figure 5.22. SAXS data obtained from 6 mg/ml BAPKO\_0422.** A) and C) are plots of the scattering. (A) shows the calculated thermal background (red), (C) has been manually adjusted to a baseline of approximately 0. B) and D) are the respective Guinier plots of (A) and (C). (B) was produced using data with the calculated thermal background subtracted. The dark blue data in plot (B) is discounted in the  $R_g$  calculation as the slope suggests slight aggregation.

The  $R_g$  was calculated for BAPKO\_0422 at 6 mg/ml assuming a spherical molecule from each Guinier plot - 18.2 Å (Figure 5.22 B) and 18.6 Å (Figure 5.22 D). The difference between the two is very small as the upturn in data at low  $q$  (Figure 5.22 B, dark blue data) which would have affected the slope of the line has been discarded from analysis. The raw data and Guinier plots for BAPKO\_0422 at 10 mg/ml are presented in Figure 5.23. The 10 mg/ml dataset was plotted in three ways (Figure 5.23); normalised using a thermal background (Figure 5.23 A), normalised to 0 by manual subtraction (Figure 5.23 C) and (C) with the data at low  $q$  removed (Figure 5.23 E). Guinier plots of each have been produced.





**Figure 5.23. SAXS data obtained from 10 mg/ml BAPKO\_0422.** A), C) and E) are plots of the raw scattering, A includes the thermal background (red line), (C) has had the data manually normalised to a baseline of 0 and truncated at high  $q$ , (E) is (C) with data truncated at low  $q$ . B), D) and F) are Guinier plots produced from (A), (C) and (E) respectively. The dark blue data points in (B) are not included in the  $R_g$  calculation, in addition, (B) has had the thermal background subtracted. E is the same dataset as (C) with the first 12 data points removed as the angle of the slope indicated aggregation.

The  $R_g$  calculated using the slope of the Guinier plot Figure 5.23 B was 22.1 Å, compared to that of 21.4 Å from Figure 5.23 D (the upturn in data at low  $q^2$  was not included). The  $R_g$  calculated from Figure 5.23 F was 21.4 Å. As with 6 mg/ml datasets, the calculated  $R_g$  from 10 mg/ml datasets are in close agreement, this is however a rough approximation as

evidenced by the 3 Å difference between the  $R_g$  values.  $R_g$  can be calculated more accurately by GNOM from the  $P(r)$  function, which encompasses the whole dataset rather than a small section, rather than from the slope of a small region of data points from a Guinier plot (Jacques and Trewhella, 2009). GNOM was used to calculate the  $R_g$  from one 6 mg/ml dataset (Figure 5.22 A) - henceforth denoted as 6TT (**6** mg/ml data, adjusted for **T**hermal background and **T**runcated - data at low  $q$  removed), and two datasets from 10 mg/ml data - Figure 5.23 C, 10N (**10** mg/ml data, manually **N**ormalised, data at low  $q$  left in) and 10NT Figure 5.23 E (**10** mg/ml data, manually **N**ormalised and **T**runcated - data at low  $q$  removed). The calculated  $R_g$  values are presented in Table 5.5.

$I(0)$  is the scattering intensity at zero-angle, although this cannot be distinguished from X-rays that pass through the sample unscattered; this value can however be extrapolated (Jacques and Trewhella, 2009).  $I(0)$  is proportional to sample concentration, and inconsistencies can indicate imperfections in the sample. A decrease in the  $I(0)$ :concentration ratio with an increase in concentration suggests interparticle interference (generally observed in conjunction with a decrease in  $R_g$ ). An increase in  $I(0)$ :concentration ratio indicates concentration dependent aggregation or oligomerisation.

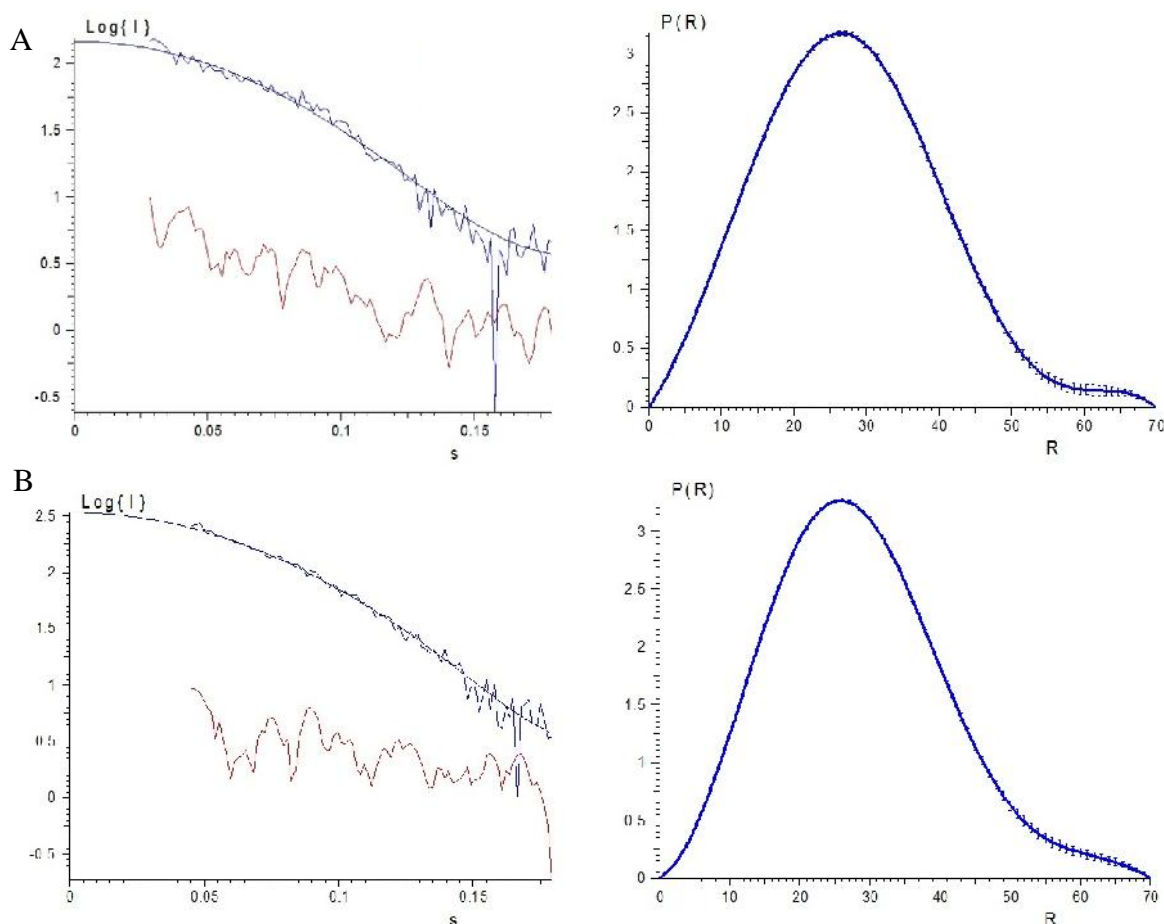
Program	Parameter	6TT	10N	10NT
MANUAL	<b>Rg (Å)</b>	18.6	21.4	21.4
	<b>Rg</b>	17.7 ± 0.03	21.8 Å ± 0.04	23.1 Å ± 0
	<b>sRg limits</b>	0.857 to 1.21	1.177 to 1.704	1.152 to 1.58
PRIMUS	<b>I(0)</b>	116.86	338.94	368.41
	<b>Quality</b>	76 %	24 %	41 %
	<b>Notes</b>	-	<b>AGGREGATED</b>	-
GNOM	<b>Highest ALPHA</b>	0.3796E+01	0.4045E+02	0.1449E+02
	<b>Total estimate</b>	0.617	0.452	0.881
	<b>Rated</b>	<b>REASONABLE</b>	<b>REASONABLE</b>	<b>GOOD</b>
	<b>Reciprocal space Rg (Å)</b>	21.50	27.91	21.81
	<b>I(0)</b>	0.1460E+03	0.5694E+03	0.3412E+03
	<b>Real space Rg (Å)</b>	<b>21.50 ± 0.247</b>	<b>27.85 ± 0.032</b>	<b>21.81 ± 0.215</b>
	<b>I(0)</b>	0.1460E+03 ± 0.1890E+01	0.5693E+03 ± 0.1742E+01	0.3412E+03 ± 0.3763E+01

**Table 5.5. PRIMUS and GNOM analysis.** Included in this table are the various Rg values and parameters computed by PRIMUS and GNOM for each of the three datasets: 6 mg/ml normalised and truncated (6TT), 10 mg/ml normalised (10N) and 10 mg/ml normalised and truncated (10NT). The calculated Rg values differ between PRIMUS and GNOM due to the methods of calculation. Comparing the real space Rg and  $I(0)$  values of datasets 6TT and 10NT calculated by GNOM; the real space Rg value remains constant with an increase in concentration, however the  $I(0)$  value more than doubles. An increase in these values is indicative of aggregation, however as only the  $I(0)$  increases, aggregation is assumed to not be too high. The 10NT data was not processed further as the data suggested aggregation.

The Rg for the 10N dataset increased considerably from the 6TT dataset, suggesting that this data was aggregated. However the Rg calculated from 10NT data showed good agreement with 6TT Rg. The 10N dataset was discounted from analysis due to the Rg and  $I(0)$  compared to 6 mg/ml data indicating aggregation and GNOM pronouncing the data to be “Aggregated”, indicating that 10 mg/ml is probably reaching the solubility limit of the protein.

### 5.3.2.2 Evaluation of the particle distance distribution function

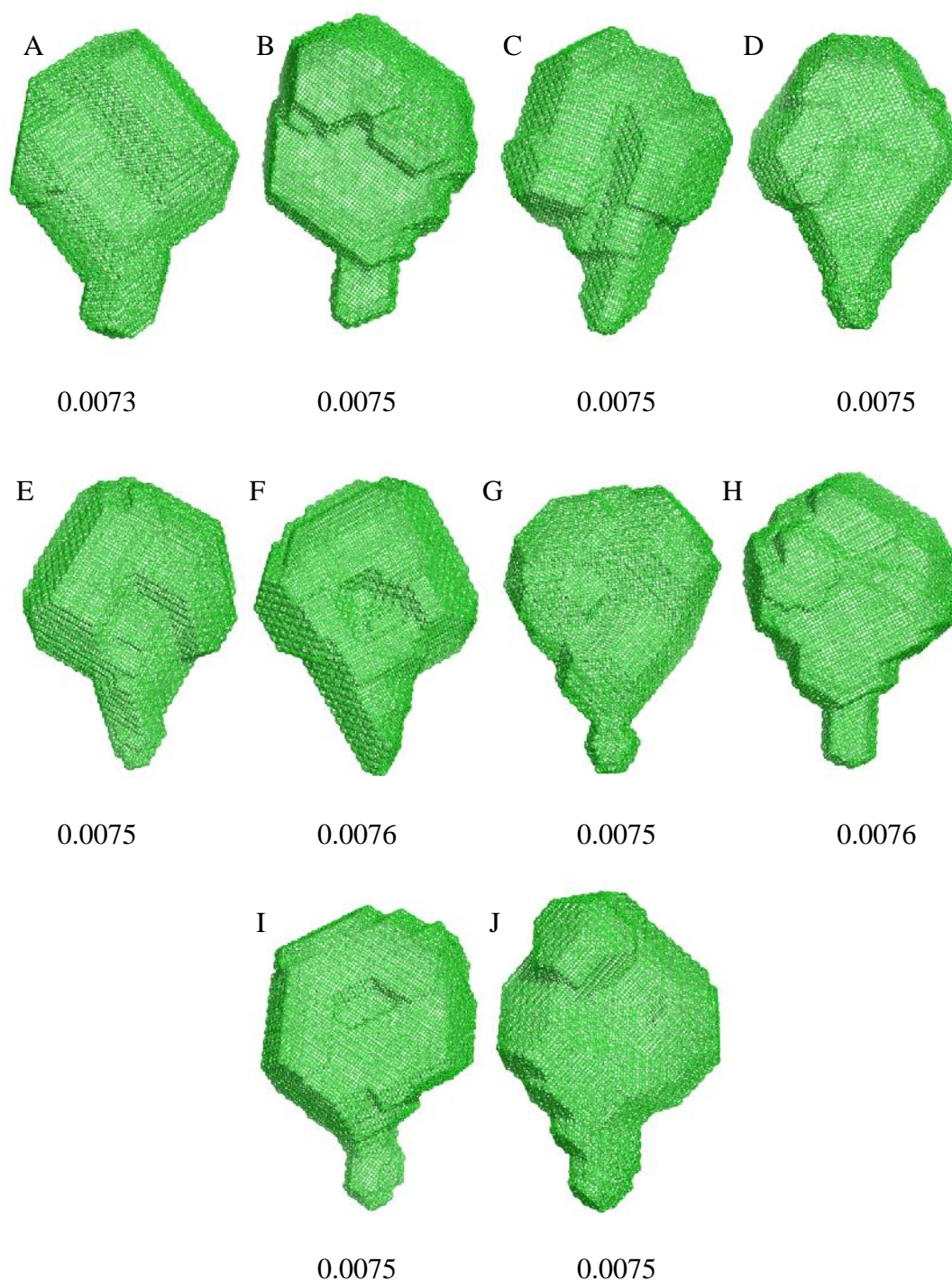
The datasets acquired at both 6 and 10 mg/ml were run through GNOM using a  $D_{\max}$  of 70 Å for calculation of  $P(r)$  (Figure 5.24). The plots do not reach 0 at a steady slope like the control lysozyme experiment. This is indicative of particle overlap potentially caused by the flexible His-tag at the N-terminus.



**Figure 5.24. GNOM fit and  $P(r)$  plots of BAPKO\_0422 datasets. A) dataset 6N; B) dataset 10NT.** The smooth blue line (Ireg) is the GNOM fit to the data, the other blue line is the experimental data (Jexp) and the red line is the experimental input points used for regularisation (Jexp) (Svergun, 1992). 'q' is designated 's' in the program. The fit of the data to predicted models is good for datasets 6NT and 10NT. The  $P(r)$  function for both suggests that even at a distance of 70 Å there is still particle overlap; this is likely due to either the N-terminal His-tag or flexible loop regions.

### 5.3.2.3 Production of molecular envelope model using dataset 6TT

The GNOM output file from the 6 mg/ml dataset (6TT) was used by DAMMIF to produce the following molecular envelopes (Figure 5.25).



**Figure 5.25. Molecular envelopes generated from dataset 6TT.** The fit score is included underneath each solution. All solutions have excellent scores and similar overall shape. The region of electron density at the bottom of the model could be a product of the N-terminal his-tag. The protein is predicted to be barrel shaped with flexible loop regions at either end, the overall diamond shape of the bulk of the density does not appear to match the shape, unless detergent molecules around the hydrophobic segment are taken into account, in which case it is possible that the molecular envelope solution is appropriate.

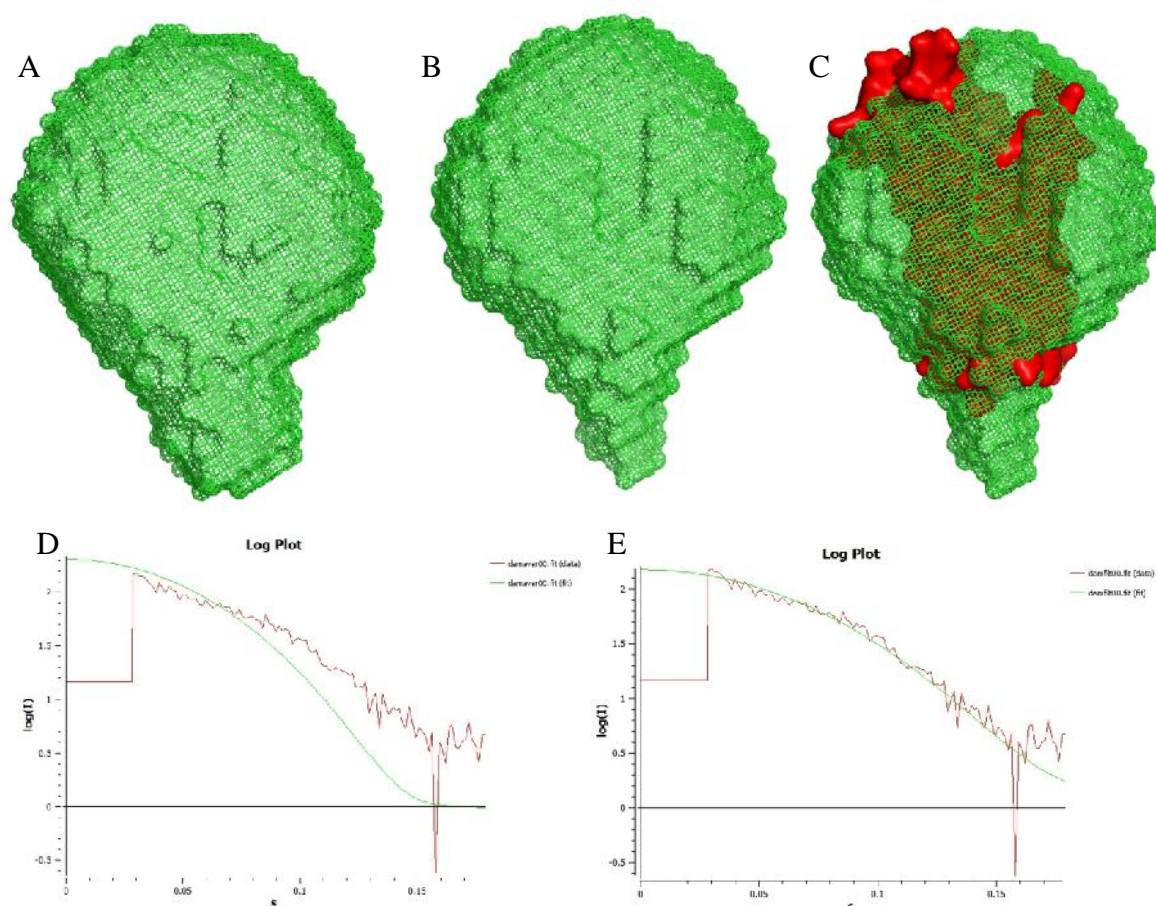
Each of the molecular envelopes (Figure 5.25) was evaluated using DAMSEL and DAMSUP, which found that all solutions were acceptable (Table 5.6).

Solution	Average DAMSEL comparison NSD	Include/discard for DAMSUP	DAMSUP superimposition NSD value
F	0.523	Reference	Reference
C	0.537	<b>Include</b>	0.437
E	0.537	<b>Include</b>	0.512
B	0.538	<b>Include</b>	0.532
A	0.539	<b>Include</b>	0.488
J	0.548	<b>Include</b>	0.542
H	0.550	<b>Include</b>	0.559
D	0.555	<b>Include</b>	0.521
I	0.564	<b>Include</b>	0.562
G	0.579	<b>Include</b>	0.550

**Table 5.6. DAMSEL comparison of 6TT molecular envelope solutions.** The overall mean NSD for all models was 0.547 and the variance 0.016; giving NSD limits for accepting models of 0.515 and 0.579. All of the models fit within this range and so all were included for further analysis. Each of the solutions shown in Figure 5.25 was accepted by DAMSEL (third column from left above) and a molecular envelope created by superimposition onto solution F. The NSD values relative to this reference were calculated.

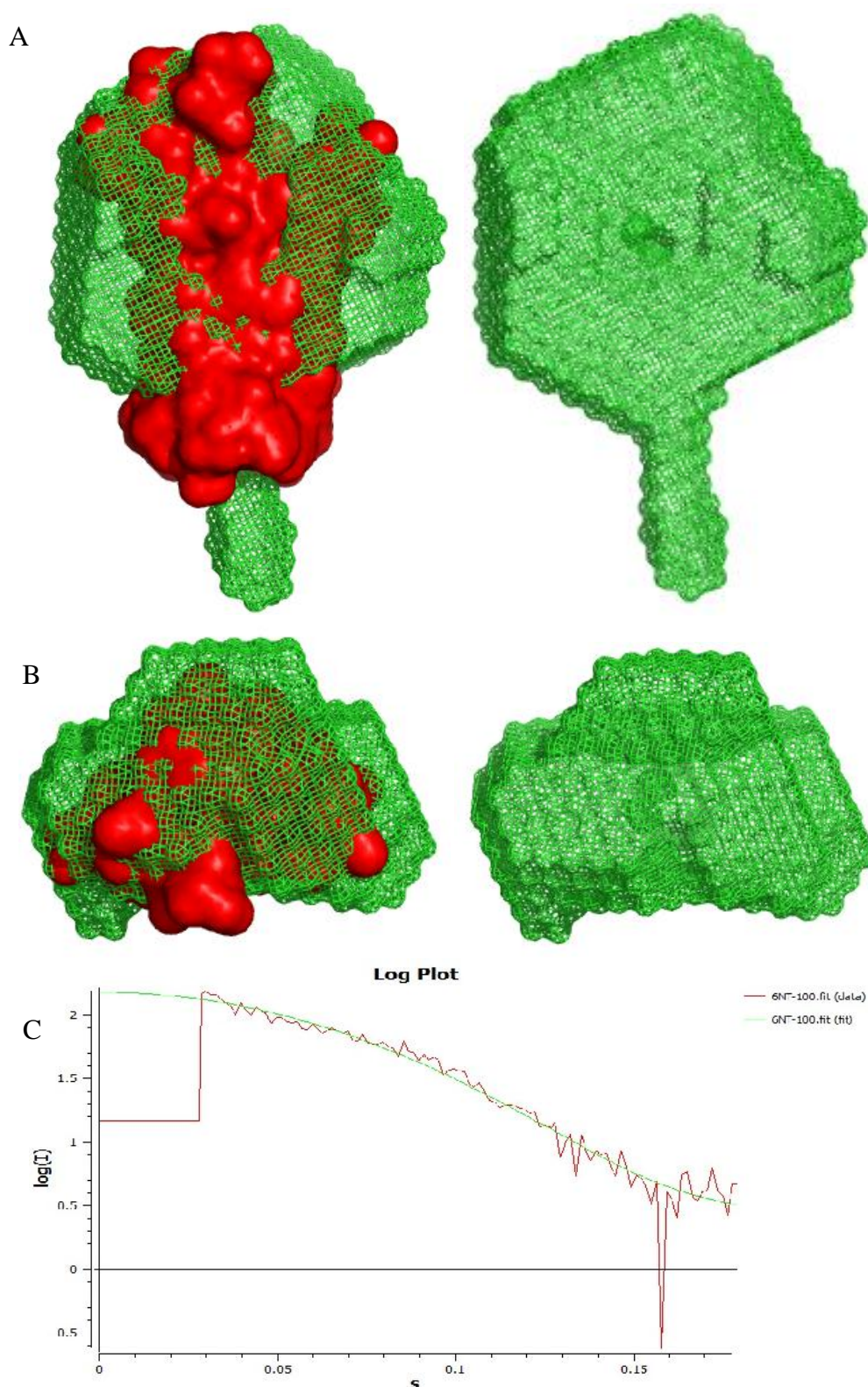


The molecular envelopes (Figure 5.25) were averaged by DAMAVER, refined using DAMFILT (Figure 5.26 A - C) and the fit to the data plotted using CRY SOL (Figure 5.26 D & E).



**Figure 5.26. 6TT molecular envelopes after refinement.** **A)** DAMAVER averaged molecular envelope; **B)** DAMFILT refined molecular envelope; **C)** DAMFILT molecular envelope with 1BXW crystal structure for comparison (coloured red for ease of visibility). DAMFILT has trimmed away density at the edges of the DAMAVER averaged solution resulting in a rounder body of electron density and a more pointed region at the bottom of the model. 1BXW fits in the generated molecular envelope relatively well, although the loop regions make the structure larger than the region of density and BAPKO\_0422 is predicted to have longer loops than 1BXW. The small region of density at the bottom of the model could be due to a slightly longer periplasmic turn, or the His-tag; although this is unlikely as the density is too small to accommodate the peptide chain. **D)** CRY SOL plot of the fit of the molecular envelope generated by DAMAVER to the experimental data; **E)** the fit of the DAMFILT solution to the experimental data. DAMAVER as expected produced a solution that does not fit the experimental data, the DAMFILT solution however is reasonable.

The DAMAVER averaged solution was used by DAMSTART to produce a model of fixed core. DAMMIN used this ‘start-file’ to produce the final molecular envelopes from the 6 mg/ml dataset (Figure 5.27 A & B). CRY SOL evaluation (Figure 5.27 C) showed an excellent overall fit of the solution to the data.

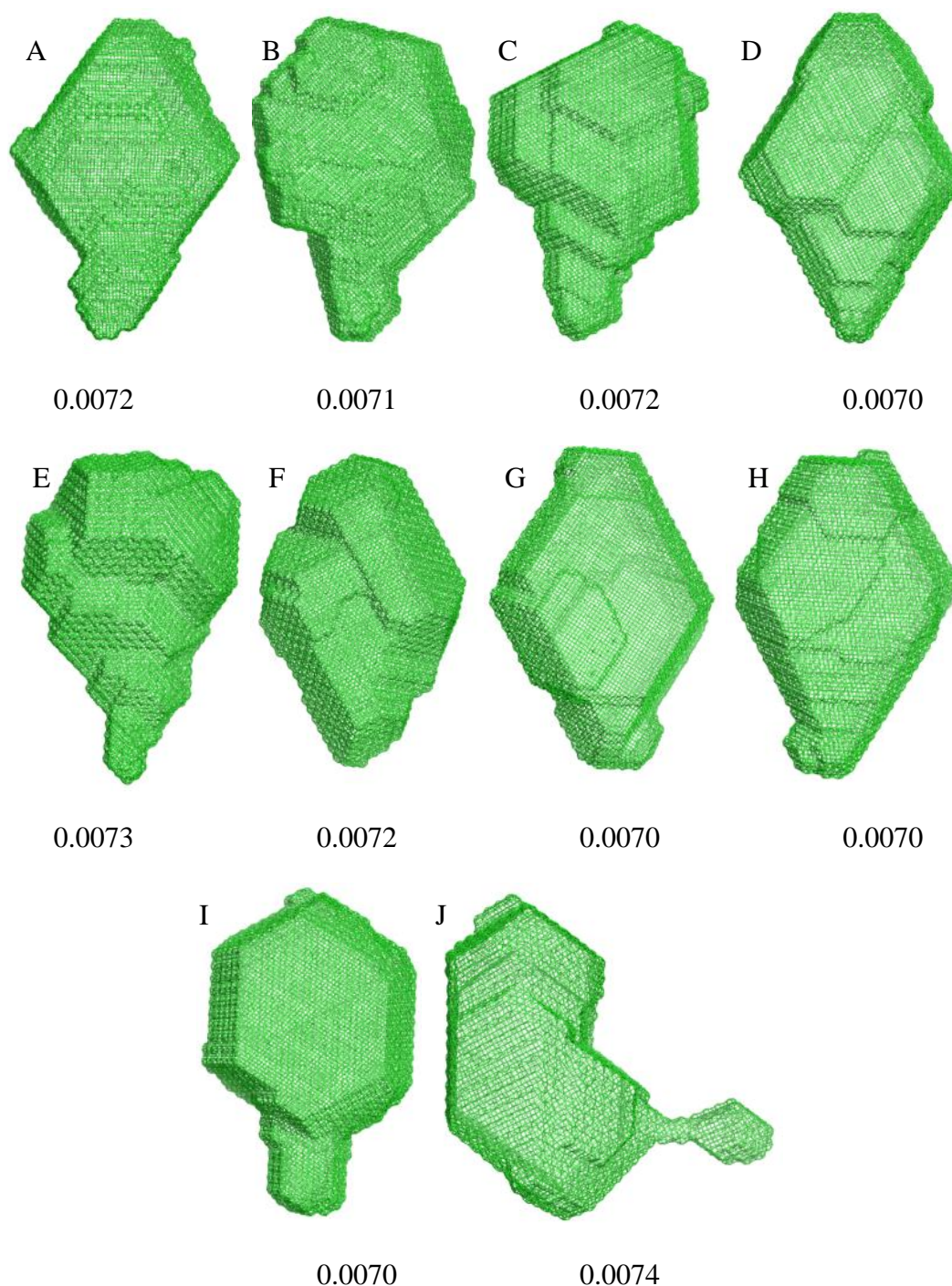


**Figure 5.27. DAMMIN refined molecular envelope 6TT.** DAMMIN refinement of the molecular envelope. Here the front **A**) and top down **B**) views are shown with and without the crystal structure of 1BXW for comparison (showing solvent accessible surface, coloured red for easier visibility). **C**) shows the fit of the model (green) to the experimental data. This was produced in CRY SOL and the produced molecular envelope fits very well, indicating a good representation of the data (ChiSQR T value 1.914). The molecular envelope appears to be too short and shallow to properly fit the structure, although there is the possibility that the barrel of BAPKO\_0422 is a ‘squashed’ barrel similar to OmpX.



#### 5.3.2.4 Production of molecular envelope using dataset 10NT

The GNOM output file from the 10 mg/ml dataset (10NT) was used to produce the following molecular envelopes (Figure 5.28).



**Figure 5.28. Molecular envelopes generated from dataset 10NT.** Analysis was repeated ten times using DAMMIF with slow annealing. The fit score is included underneath each solution, and all solutions have an excellent fit score. There appears to be two major types of solution to the data; solutions (A), (D), (F), (G) and (H) are roughly diamond shaped; whilst solutions (B), (C), (E) & (I) appear to have a region of density at the bottom (as viewed). Solution (J), despite having a good fit score is highly unlikely to be an accurate solution. The expected protein shape is a barrel with flexible loop regions and detergent molecules bound around the centre of the barrel. The diamond shaped solutions may account for this.

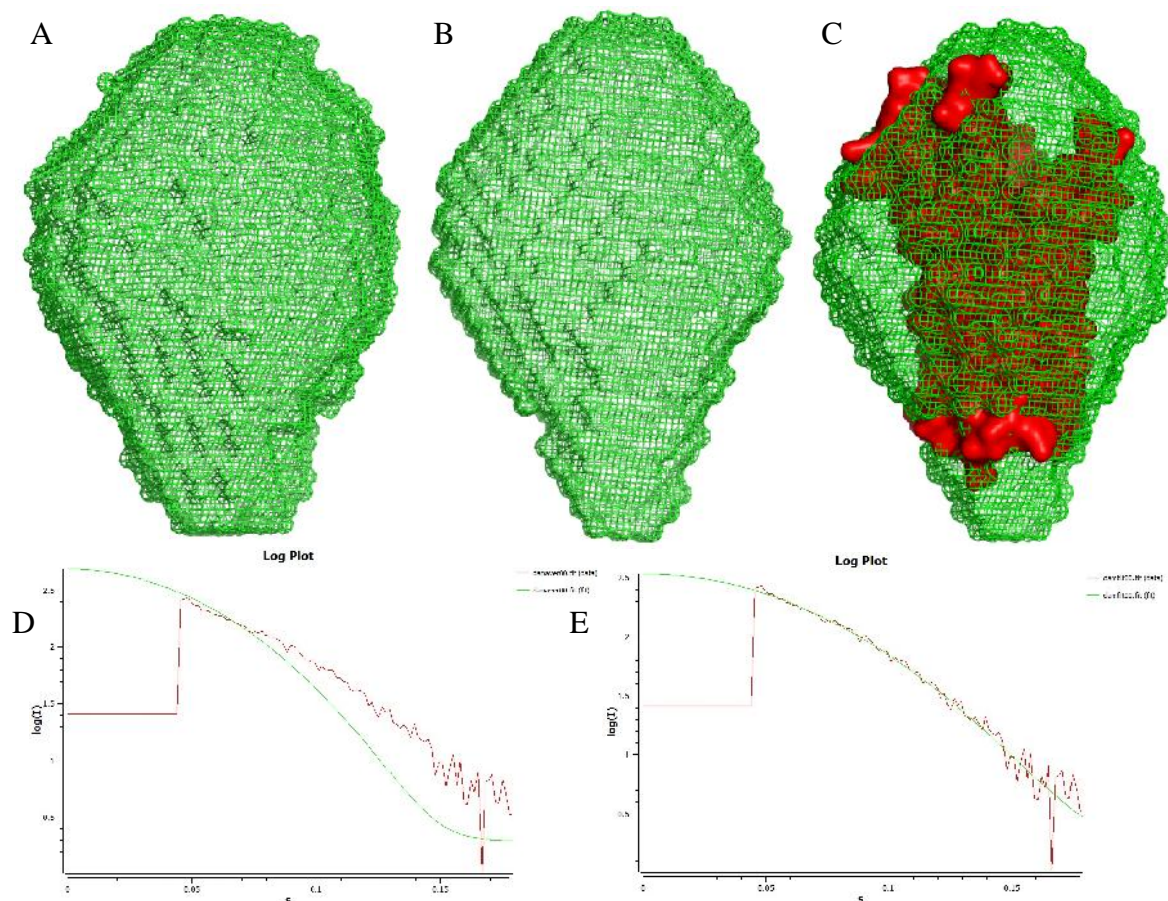
The similarity of generated solutions was compared using the DAMSEL component of DAMAVER. The DAMSEL output file was run through DAMSUP to produce molecular envelopes superimposed over the reference model and NSD values for each (Table 5.7).

Solution	Average DAMSEL comparison NSD	Include/discard for DAMSUP	DAMSUP superimposition NSD value
H	0.559	Reference	Reference
A	0.574	<b>Include</b>	0.507
B	0.574	<b>Include</b>	0.519
I	0.581	<b>Include</b>	0.494
F	0.586	<b>Include</b>	0.501
D	0.594	<b>Include</b>	0.510
G	0.595	<b>Include</b>	0.482
E	0.602	<b>Include</b>	0.575
C	0.603	<b>Include</b>	0.527
J	0.907	<b>Discard</b>	-

**Table 5.7. DAMSEL comparison of 10NT molecular envelope solutions.** The calculated mean NSD value for all ten solutions was 0.617, with a variation of 0.103. Any solutions with an NSD more than twice the variance away from the mean are discarded, in this case the NSD outer limits for accepting solutions are 0.411 and 0.823; hence solution (**J**) was discounted from following analysis. The DAMMIF generated molecular envelope models were superimposed onto the reference solution (molecular envelope (**H**) - Figure 5.28) and the NSD values relative to this calculated. Solution (**J**) was discarded as it did not fulfil the criteria in DAMSEL.

Solution J was discarded due to the difference in shape, despite having a score equivalent to those of the other solutions.

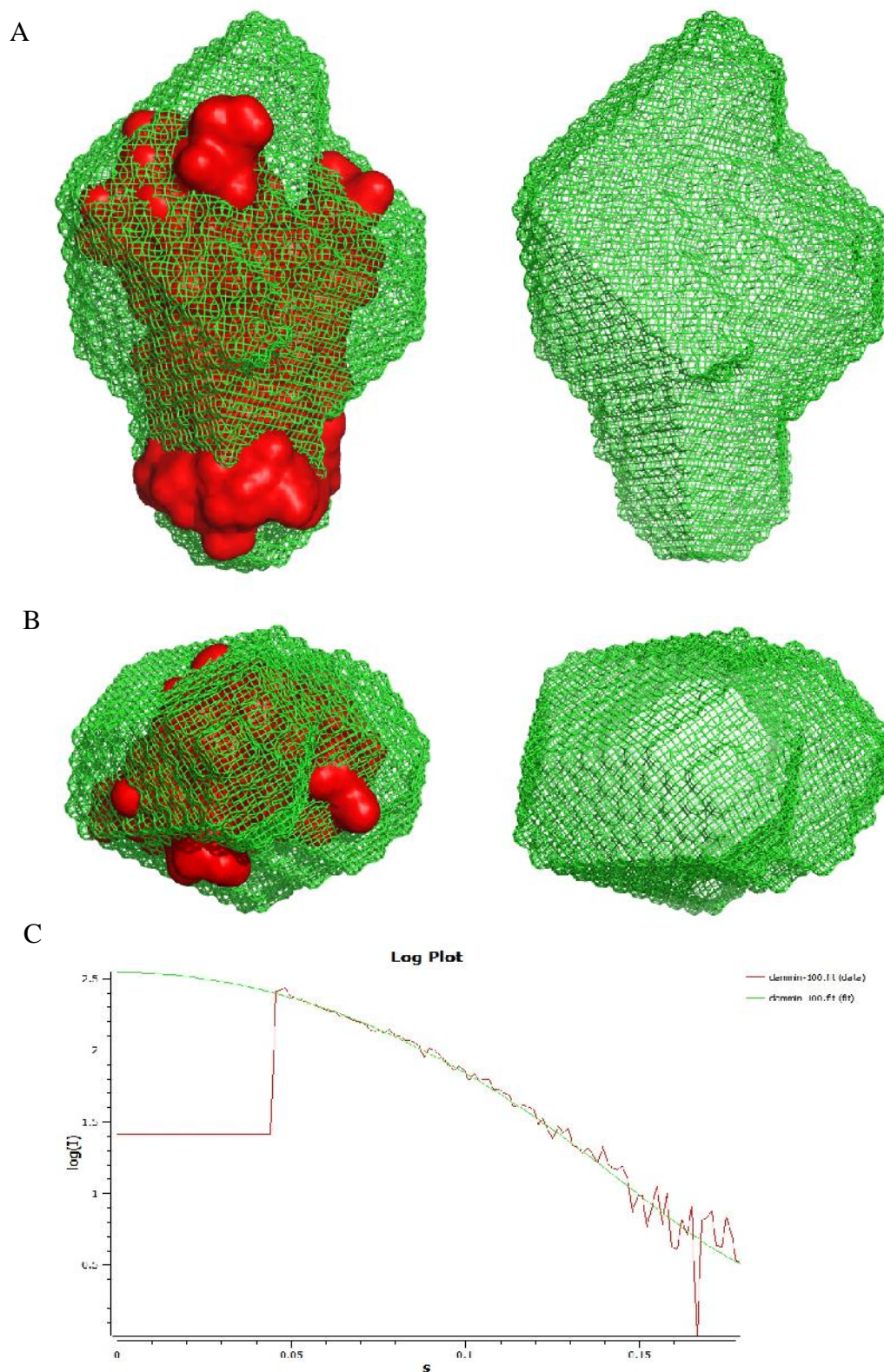
The DAMSUP output file was run through DAMAVER, which produced an averaged molecular envelope (Figure 5.29 A). This was refined using DAMFILT, which removed low occupancy atoms and produced the molecular envelope that is most probable (Figure 5.29 B & C). The fit of the models to the data was compared using CRY SOL (Figure 5.29 D & E).



**Figure 5.29. 10NT data averaged molecular envelopes.** A) DAMAVER averaged molecular envelope; B) DAMFILT refined molecular envelope; C) DAMFILT molecular envelope with 1BXW crystal structure for comparison (coloured red for ease of visibility). DAMFILT has trimmed away density at the edges of the DAMAVER averaged solution, and 1BXW fits in the generated molecular envelope well. BAPKO\_0422 is predicted to have longer loop regions than 1BXW, which would not fit within the molecular envelope calculated here. The result is still reasonable and fits the data very well. D) shows a very poor fit of the DAMAVER envelope to the data; E) highlights a good fit of the DAMFILT solution to the data. Both (D) and (E) produced in CRY SOL (data in red, fit in green.).

It should be noted that the DAMFILT molecular envelope is not expected to be a good fit for the experimental data. As this is the case, the model was prepared for refinement in DAMMIN using DAMSTART. This generated a fixed a core of atoms from the DAMAVER model and is approximately a half the expected volume of the particle. The experimental data and the present solution was run through DAMMIN to produce a molecular envelope that should fit the experimental data (Figure 5.30 A & B). CRY SOL analysis showed an excellent overall fit to the data (Figure 5.30 C).



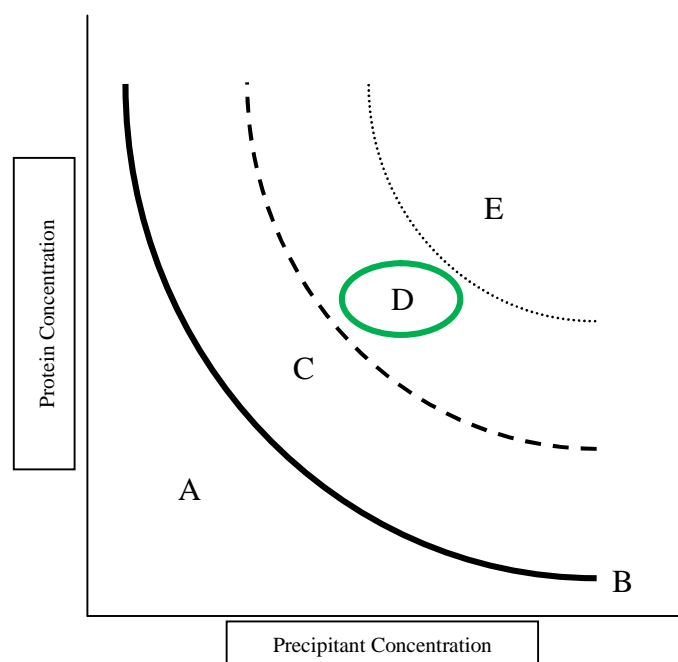


**Figure 5.30. DAMMIN refined molecular envelope 10NT.** The DAMFILT produced molecular envelope was run through DAMMIN to further refine the molecular envelope. Here the front (**A**) and top down (**B**) views are shown with and without the solvent accessible surface of the crystal structure of 1BXW for comparison (coloured red for easier visibility). **C**) shows the fit of the model (green) to the experimental data (ChiSQRT value 1.502). This was produced in CRY SOL and the produced molecular envelope fits very well, indicating a good representation of the data.

## 5.4 Crystallisation trials of BAPKO\_0422

To solve the structure of proteins using X-ray diffraction patterns the growth of ordered protein crystals is required. This occurs when the protein molecules adopt either one or more (generally not more than two) identical orientations and interact non-covalently in the process of solidifying (Rhodes, 2000). The orientations of, and interactions between are repeated throughout the crystal. The smallest repeating 'pattern' is known as the unit cell.

Crystallisation of proteins can be achieved using various methods; traditionally through the slow, controlled precipitation of protein molecules in a 'mother-liqueur'; non-denaturing, aqueous conditions including organic and ionic compounds and polymers such as polyethylene glycol (PEG) (Rhodes, 2000). The gradual evaporation of water from the drop decreases the total volume, increasing the concentration of protein molecules, forcing them closer together; this supersaturates the drop with protein and enabling spontaneous nucleation to occur Figure 5.31. Rapid precipitation of protein samples results in unordered protein crystals from which no data can be obtained (Rhodes, 2000).



**Figure 5.31. Theoretical protein crystallisation phase diagram.** For a protein crystal to grow, the protein molecules need to supersaturate the drop to the point where spontaneous nucleation can occur (circled green). **A)** Protein undersaturated, hence the drop should remain clear. **B)** Solubility limit reached when the protein is in equilibrium with precipitant. **C)** Metastable zone, supersaturation of protein is high enough to support growth of protein crystals, but not high enough to support spontaneous nucleation. If a form of seeding is being used, this is the ideal zone. Drops in this zone are also clear. **D)** Labile supersaturation zone, spontaneous nucleation of protein molecules can occur in this zone. **E)** Highly supersaturated precipitation zone. Protein is present in too high concentration leading to amorphous precipitate. Figure adapted from (Bergfors, 2003).

The formation of protein crystals is a trial and error process, not only dependent on the concentrations of protein and precipitant, but also on incubation temperature, mother liquor pH, age of the protein, and ionic strength of the precipitation. Optimisation of the conditions to improve either the crystal size or diffraction is conducted by inducing slight variation in these factors, which can either enhance or hinder crystal formation; in addition impurities in the sample and high concentrations of precipitant can either prevent crystal formation or result in precipitation out of solution (Rhodes, 2000).

Various method exists for enhancing crystal quality besides optimisation of precipitants. Nucleation points for crystal growth can be created through the use of crystals broken down into small pieces (macroseeding) or submicroscopic pieces (microseeding); strands of hair and grains of sand have also been used to provide focal nucleation points (Bergfors, 2003). Some proteins require the addition of ligands to assist in crystallisation; these can include detergents in membrane proteins, or heavy metal cofactors; this can also be used to determine the interactions of proteins with said ligands or metals. Ligands and cofactors can also be applied to the protein crystals by ‘soaking’ (Rhodes, 2000).

#### 5.4.1 Known crystallisation of TM $\alpha$ -barrel proteins from *E. coli*

The structures of N-terminal *E. coli* OmpA, and *E. coli* OmpW and OmpX have been solved by X-ray crystallography (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Vogt and Schulz, 1999, Pautsch and Schulz, 2000, Hong *et al.*, 2006). Conditions used to grow these crystals are included in Table 5.8. These conditions may provide a guideline for protein concentrations and crystallisation conditions.

	OmpA	OmpX	OmpW	
Detergent	0.6% (w/v) C <sub>8</sub> E <sub>4</sub>	0.6% (w/v) C <sub>8</sub> E <sub>4</sub>	0.1% (w/v) LDAO	
Concentration mg/ml	20	20	9	
pH	4.5	4.6	8.5	9
Salt	1.4 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	-	0.2 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	0.3 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
Buffer	0.1 M ammonium acetate	0.1 M acetate	0.1 M Tris	0.1 M Bis-Tris propane
Additives	3.8 % (v/v) 2-propanol		25 % w/v PEG 3350	17 % w/v PEG 3350
	1.8 % (w/v) hexyldimethylaminoxide	58 % (v/v) methyl-pentanediol		
	0.5 % (w/v) C <sub>8</sub> E <sub>4</sub>			

**Table 5.8. Known crystallisation conditions of  $\alpha$ -barrel proteins.** Data acquired from (Pautsch and Schulz, 1998, Pautsch *et al.*, 1999, Vogt and Schulz, 1999, Pautsch and Schulz, 2000, Hong *et al.*, 2006).

### 5.4.2 Vapour diffusion crystallisation trials

Refolded BAPKO\_0422 in 0.1 % (w/v) LDAO was concentrated to 10 mg/ml (as determined by Bradford Assay) through the use of a 10,000 mwco centrifugal concentrator. Crystal screening was carried out using 24 condition Clear Strategy Screens (CSS) screens I and II (made in house, buffered at pH 4.5 and 8) and a 96 condition Memb-PASS Crystallisation Screen (Jena Bioscience, Jena, Germany). Different drop sizes were used (1 and 2  $\mu$ l) consisting of protein and well solution in a ratio of 50:50 and the trays incubated at 20 °C.

Eight CSS crystal screens were set up; CSS I at pH 4.5 and pH 8, and CSS II and pH 4.5 and pH 8. These were done in duplicate, at protein concentrations of 2 and 10 mg/ml. The Memb-pass screen was set up in four 24 well plates at 10 mg/ml. All screens were observed under the microscope immediately after set up and weekly thereafter for any precipitates, aggregations, phase separations and crystals. For CSS screens I and II the well solution consisted of 0.9 ml precipitant and 0.1 ml pH buffer. 0.4 ml Memb-PASS well solution was used neat (no pH buffer required). The conditions in the CSS screens are included (Table 5.9 & Table 5.10), conditions in the Memb-Pass screen are included in the Appendix 8.3.

CSS I						
	1	2	3	4	5	6
A	0.3M NaAc 25 % (w/v) PEG 2Kmme	0.2M Li <sub>2</sub> SO <sub>4</sub> 25 % (w/v) PEG 2Kmme	0.2M MgCL <sub>2</sub> 25 % (w/v) PEG 2Kmme	0.2M KBr 25 % (w/v) PEG 2Kmme	0.2M KSCN 25 % (w/v) PEG 2Kmme	0.8M NAFormate 25 % (w/v) PEG 2Kmme
B	0.3M NaAc 15 % (w/v) PEG 4k	0.2M Li <sub>2</sub> SO <sub>4</sub> 15 % (w/v) PEG 4k	0.2M MgCL <sub>2</sub> 15 % (w/v) PEG 4k	0.2M KBr 15 % (w/v) PEG 4k	0.2M KSCN 15 % (w/v) PEG 4k	0.8M NAFormate 15 % (w/v) PEG 4k
C	0.3M NaAc 10 % (w/v) PEG 8K 10 % (w/v) PEG 1K	0.2M Li <sub>2</sub> SO <sub>4</sub> 10 % (w/v) PEG 8K 10 % (w/v)m PEG 1K	0.2M MgCL <sub>2</sub> 10 % (w/v) PEG 8K 10 % (w/v)m PEG 1K	0.2M KBr 10 % (w/v) PEG 8K 10 % (w/v)m PEG 1K	0.2M KSCN 10 % (w/v) PEG 8K 10 % (w/v)m PEG 1K	0.8M NAFormate 10 % (w/v) PEG 8K 10 % (w/v)m PEG 1K
D	0.3M NaAc 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme	0.2M Li <sub>2</sub> SO <sub>4</sub> 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme	0.2M MgCL <sub>2</sub> 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme	0.2M KBr 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme	0.2M KSCN 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme	0.8M NAFormate 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme

**Table 5.9. CSS I screen.** Protein screen that is optimised to minimise the production of salt crystals whilst screening a broad range of precipitants. The screen is used in conjunction with CSS II at two pHs; Tris at pH 8 and Sodium Acetate at pH 4.5.

CSS II						
	1	2	3	4	5	6
<b>A</b>	1.5M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	0.8M Li <sub>2</sub> SO <sub>4</sub>	2M NaFormate	0.7M KH <sub>2</sub> PO <sub>4</sub>	0.2M CaAcetate 25 % (w/v) 2Kmm	0.2M CaAcetate 15 % (w/v) 4K
<b>B</b>	2.7M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.8M Li <sub>2</sub> SO <sub>4</sub>	4M NaFormate	1M KH <sub>2</sub> PO <sub>4</sub>	0.2M CaAcetate 10 % (w/v) PEG 8K 10 % (w/v) PEG 1K	0.2M CaAcetate 8 % (w/v) PEG 20K 8 % (w/v) PEG 550mme
<b>C</b>	40 % (w/v) MPD	40 % (w/v) butane- 1,4-diol	10mM CdCl <sub>2</sub> 20 % (w/v) 4k	0.15M KSCN 20 % (w/v) 550mme	0.15M KSCN 20 % (w/v) 600	0.15M KSCN 20 % (w/v) PEG 1.5K
<b>D</b>	35 % (w/v) Isopropanol	-	10mM NiCl <sub>2</sub> 20 % (w/v) PEG 4k	0.15M KSCN 18 % (w/v) PEG 3350	0.15M KSCN 18 % (w/v) PEG 4K	0.15M KSCN 15 % (w/v) PEG 6K

**Table 5.10. CSS II screen.** 24-well protein screen used in conjunction with CSS I screen, this screen is designed to cover other commonly used precipitants. Jeffamine was not available in the department, so well D2 was set up using the the pH buffer only. The screen is used with two different pH buffers; Tris at pH 8 and Sodium Acetate at pH 4.5.

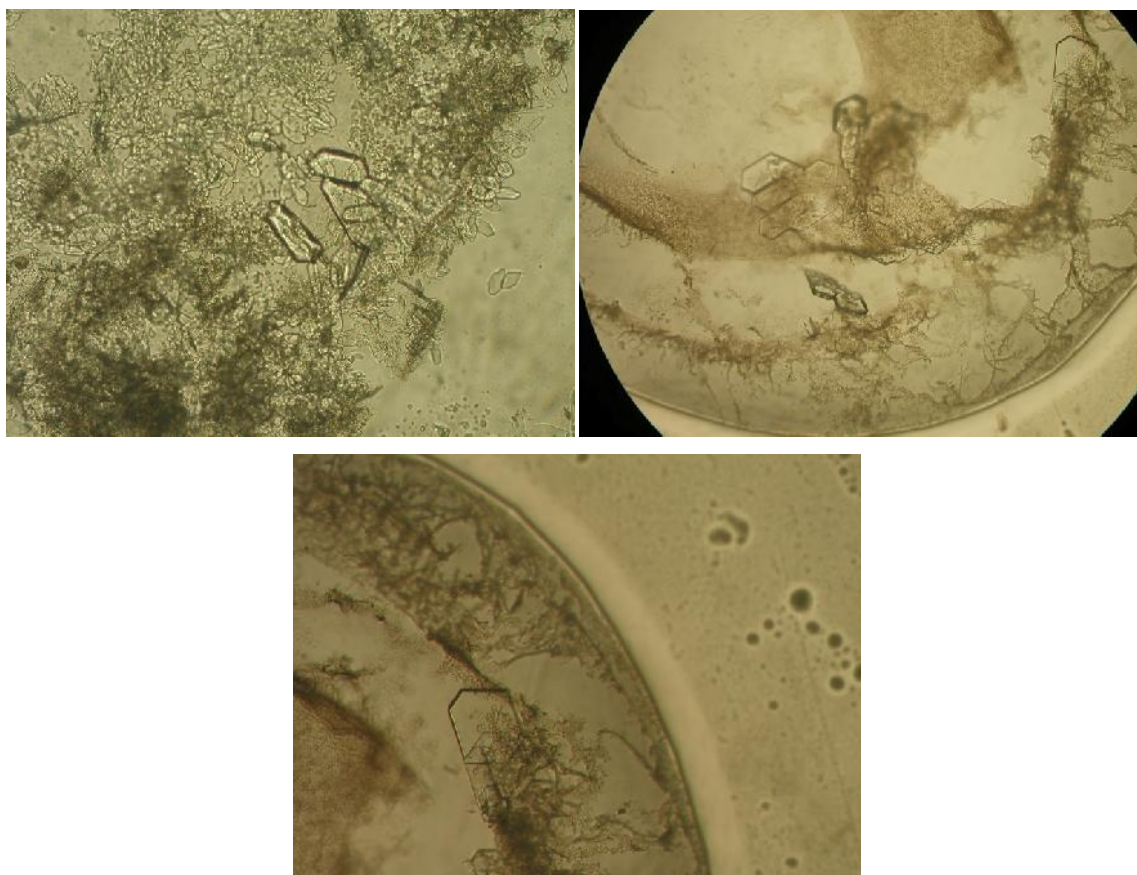
The Memb-pass screen uses a range of polyethylene glycol (PEG) solutions and differing concentrations throughout. Full details of all Memb-pass conditions can be found in appendix 8.3.

Two weeks after setup crystals were found (Figure 5.32) in the following conditions (Table 5.11).

Screen	Well	pH	Drop size (µl)	Protein concentration (mg/ml)	Conditions	Crystals
CSS I	A5	8	1	2	0.2 M KSCN, 25 % (w/v) PEG 2Kmm	Three small crystals, multiple microcrystals
CSS I	D5	4.5	1	10	0.2 M KSCN, 8 % (w/v) PEG 20K, 8 % (w/v) PEG 550mme	Three thin crystal plates, Two roughly square crystals

**Table 5.11. Results of initial crystal screening.** All other conditions investigated resulted in either a clear drop, some protein aggregation or varying levels of protein precipitating out of solution.

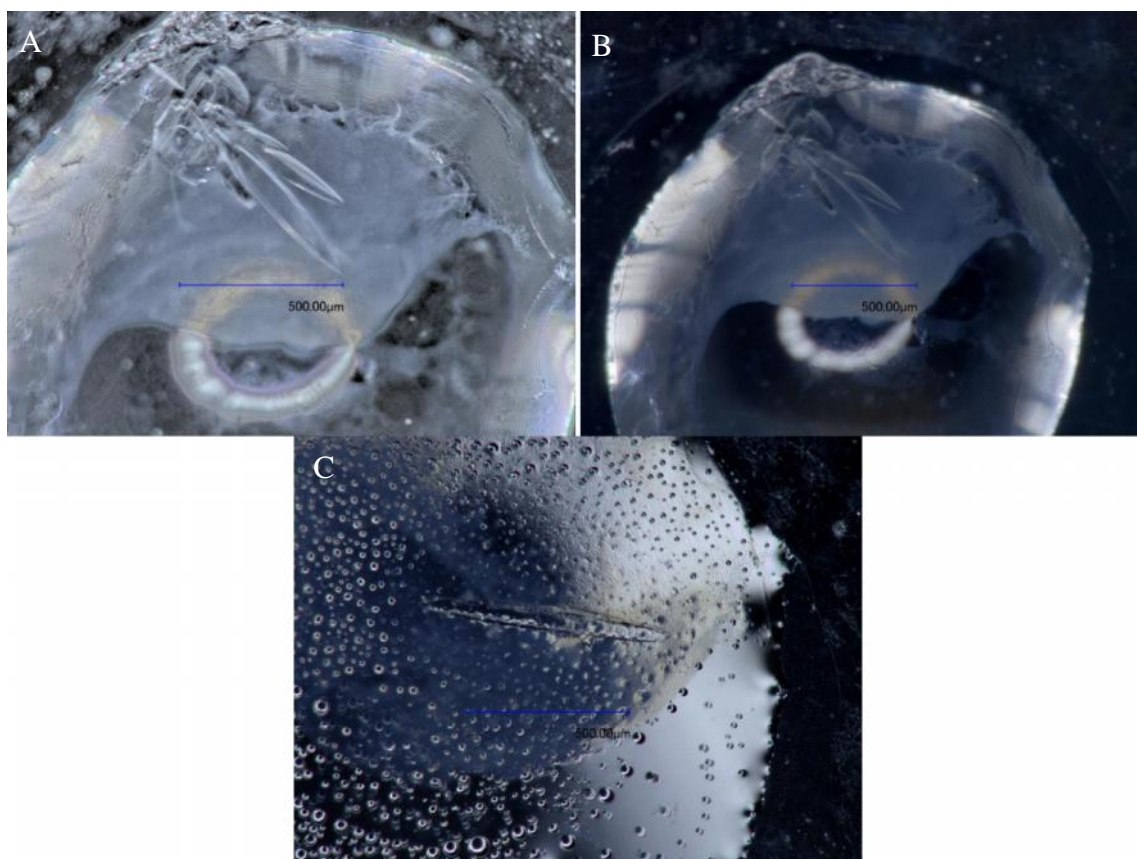




**Figure 5.32. Salt crystals obtained from CSS I and II.** Crystals were obtained from CSS I condition A5 at pH 8 and CSS I condition D5 at pH 4.5. KSCN was present in both conditions.

The crystals were frozen in well solution with the addition of 20 % (w/v) glycerol and X-ray images collected at the University of Leeds. The crystals were found to be salt crystals (results not shown).

Both CSS screens and the Memb-PASS screen were repeated at protein concentration of 5 mg/ml in 0.1 % (w/v) LDAO and the trays checked every weeks. After several months crystals were found in several Memb-PASS conditions (Figure 5.33).



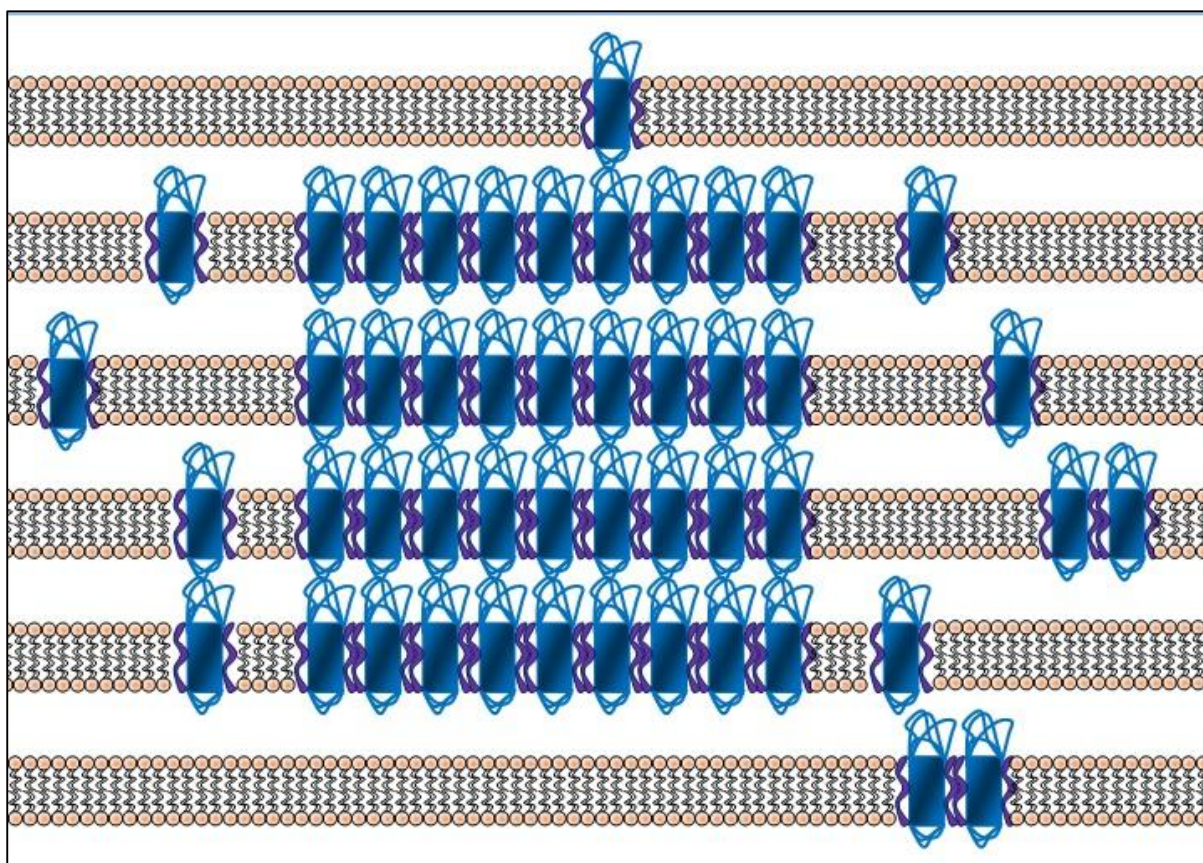
**Figure 5.33. Crystals grown by vapour diffusion.** A) & B) pointed elliptical shaped crystals growing from a single point in precipitant A1 containing 48 % (v/v) PEG 400, 100 mM sodium citrate, 210 mM sodium acetate and 1.12 % (v/v) glycerol, pH 4.5; C) elliptical shaped crystal grown in precipitant B5 – 38 % (v/v) PEG 550 MME, 100 mM ADA, 140 mM Sodium Bromide, pH 6.5.

These were also found to be salt crystals.

### 5.4.3 *In meso* crystallography

Lipid cubic phase (LCP) crystallography uses a complex mixture of hydrated monoolein (MO), water, protein, detergent, and precipitants to grow membrane protein crystals (Ai and Caffrey, 2000). Hydrated MO forms a variety of phases dependent upon the level of hydration. The phase from which protein crystals are produced is known as the cubic mesophase, hence LCP can also be referred to as *in meso* crystallography (Ai and Caffrey, 2000).

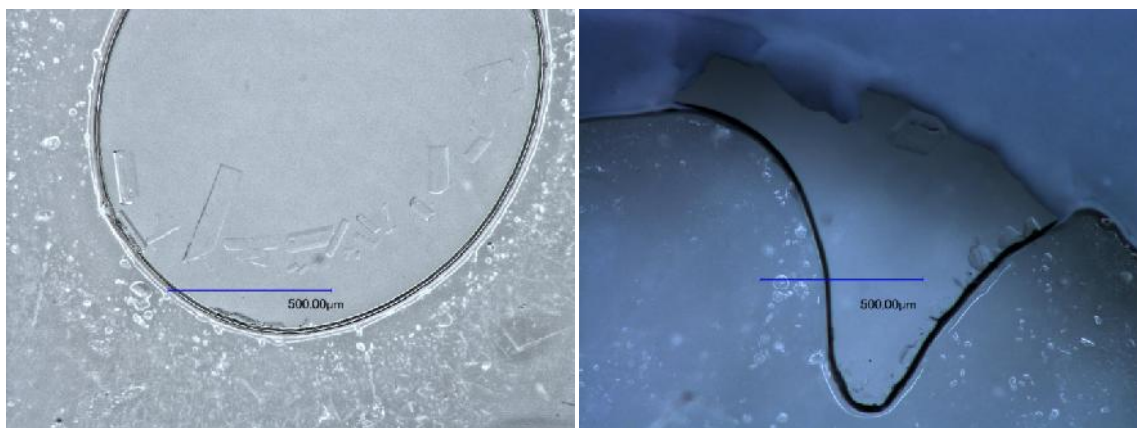
The cubic phase produced is essential to crystal growth, therefore any precipitants and/or additives that are introduced must not destabilise the cubic phase (Ai and Caffrey, 2000). Detergents that are stabilising the protein are likely to be incorporated into the cubic phase along with the protein (Ai and Caffrey, 2000). The cubic phase consists of bicontinuous curves (Nollert, 2005) and has been shown to aid in the formation of membrane protein crystals, described in more detail below (Figure 5.34).



**Figure 5.34. Formation of crystals in the lipid cubic phase.** The events purported to take place within *in meso* crystallisation begin with the protein (blue molecules) being reconstituted into the cubic phase (orange). The protein molecules diffuse into the lattice of increasing crystal phase (centre of the figure) through the sheet-like (or cubic) phase, sometimes retaining a small number of detergent molecules (purple). Once congregated, the protein molecules may form layered stacks or a regular array through the coalescing of detergent/lipid molecules, forming polar contacts between (in this case) loop regions. This process is triggered by phase separation, which in turn is a result of the addition of precipitant, upon which it is energetically favourable for protein molecules to cluster together in flattened regions of the otherwise highly curved cubic phase membrane (Nollert, 2005). The protein crystal grows as molecules are fed from a ‘reservoir’ of protein molecules in the cubic phase (Cherezov *et al.*, 2006). The protein is represented as a blue molecule, the detergent is purple and the lipid comprises small orange ellipses with a tail. Figure from (Caffrey and Cherezov, 2012).

LCP crystallisation trials were conducted using monoolein to form a lipid cubic phase to with the aim of enhancing crystallisation. LCP was conducted as described in section 2.2.17.4. The majority of drops had dried out, however in one drop (condition A9) there was a significant number of crystals (Figure 5.35).

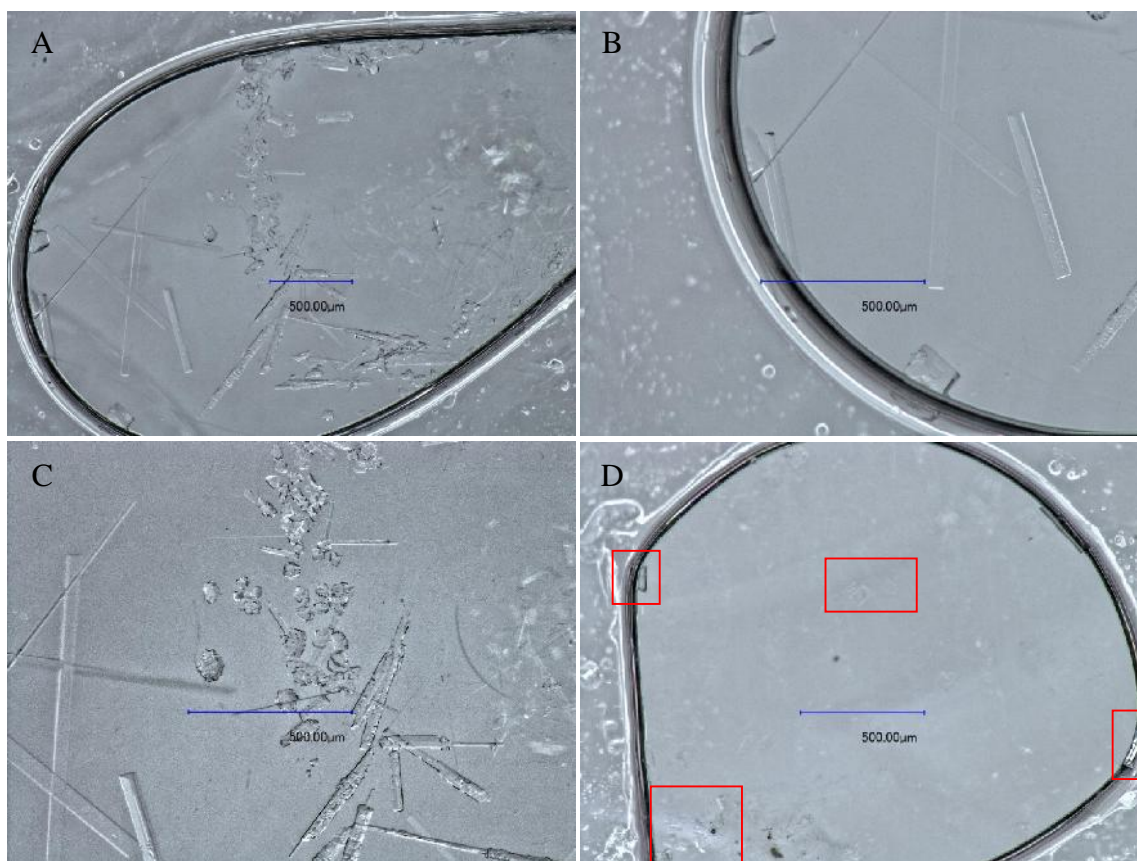




**Figure 5.35. Crystals produced using *In meso* crystallisation.** Both images are taken from the same drop. Memb-PASS condition A9; 47 % (w/v) PEG 400, 170 mM ammonium sulphate with hydrated MO. The phase resulting from mixing of MO with protein solution was too viscous to pipette, hence the phase was divided as best as possible. Once precipitant was added the drop did not remain in the centre of the well, splitting into two, one in the centre and one at the edge. Several crystals were produced in both although too small and thin to test-shoot and so an attempt was made to stain the protein crystals with methylene blue, upon which the crystals promptly dissolved. This, along with the fact that it was highly unlikely that the precipitant solution would form crystals indicates that the crystals were indeed protein.

Methylene Blue (0.005 % w/v) was added to the drop to stain the crystals, at which point the crystals rapidly dissolved.

Further lipid cubic phase crystallisation trials used wells that were formed using a grid of grease on the microscope slide. This technique was adjusted and repeated in larger volumes. The LCP crystallography repeat crystallisation used MO (~60 % w/v) hydrated with protein solution (~10 mg/ml) in a 3:2 ratio (section 2.2.15.3) with 1 µl A9 precipitant added and sealed with a coverslip. This was done in triplicate and after two days, crystals were present in two of the drops (Figure 5.36).



**Figure 5.36. Crystals grown from *In meso* crystallography.** All drops were set up using precipitant A9. Images **(A)**, **(B)** and **(C)** are from the same drop. **(B)** and **(C)** were taken at an increased magnification. **(A)** contains large thin rectangular crystals to the left of the drop; smaller unordered crystals in the centre of the drop and at the bottom and top right; and what appears to be a hexagonal phase formed from the MO on the right side of the drop.

The crystals were plate-like and difficult to manipulate. In order to differentiate between salt and protein they were stained using methylene blue. Unfortunately the stain dissolved the crystals. Crystal growth using *In meso* crystallography was shown to be highly reproducible.

## 5.5 Discussion

Circular dichroism is a commonly used technique used to determine the secondary structure composition of folded proteins in solution. CD is a non-destructive technique (provided that the protein is not heated, or is, but is thermally stable) that is capable of using a small volume of low concentration (less than  $\sim 1$  mg/ml) protein. The JASCO J-810 spectropolarimeter used was equipped with a Peltier device enabling rapid heating of samples for observation of protein unfolding. While CD is capable of gathering information at a wide range of wavelengths, this is dependent on the composition of buffers used. Impurities, along with high salt, detergent or protein concentrations can rapidly increase the H[T] voltage above the threshold (which varies from machine to machine), beyond which the data becomes unreliable and unusable. The interpretation of the data obtained is subjective to the algorithm and reference datasets used; the reference datasets are based predominantly on globular proteins with membrane spanning proteins being significantly under represented. Table 5.1 shows slight differences in data analysed using the algorithms SELCON3 and CDSSTR with two different datasets (4 and 7), Table 5.2 show slight differences in data analysed using the CDSSTR algorithm and the same two reference sets. CDSSTR was selected over SELCON3 due to CDSSTR selectively discounting proteins from the reference sets that are not relevant to the experimental protein. As  $\alpha$ -barrel membrane proteins are underrepresented in the reference sets, CDSSTR will therefore result in a more accurate calculation than SELCON3. Calculation of the secondary structure from CD spectra of BAPKO\_0422 at various concentrations and temperatures was carried out using the online webserver DichroWeb (Lobley *et al.*, 2002, Whitmore and Wallace, 2004, Whitmore and Wallace, 2008). Data was acquired down to 195 nm before the H[T] voltage exceeded the acceptable threshold of 600 V, this was potentially due to the concentration of detergent and salts in the buffer. It would have been advantageous to have repeated data acquisition in the buffer used for SAXS that contained 0.3 M NaCl, 3 mM Tris and 0.1 % (w/v) LDAO.

The known CD spectra of the N-terminal domain of *E. coli* (Figure 5.4) is very similar in shape to that obtained from BAPKO\_0422 (Figure 5.5) To obtain an initial estimate of the expected CD spectra for OmpA-like proteins the .pdb files for 1BXW, 2F1T and 1QJ8 were submitted to DichroCalc (Bulheller and Hirst, 2009) for calculation of the theoretical curve. The locations of the CD signal peaks (positive peak at approx. 196 nm and negative peak at  $\sim$

220 nm) of OmpA, OmpW and OmpX from the theoretical calculated data are very close to the positions expected (Greenfield, 2006).

The % secondary structure was calculated (Table 5.1) to be 40 %  $\alpha$ -strand, 20 %  $\alpha$ -turns and 33 % unordered. This is approximately what was expected for a TM  $\alpha$ -barrel protein, however the H[T] value exceeded the acceptable threshold at around 198 nm. The algorithm CDSSTR and reference sets 4 & 7 are designed for data within the range 195 - 260 nm, hence the concentration was lowered, allowing acquisition of data to a wavelength of 195.5 nm. Potential causes of the early surpassing of the upper threshold are likely to be the purity of buffer components and / or the concentration of detergent.

Heating the protein sample (0.166 mg/ml) to five different temperatures including 20 °C (Figure 5.9) yielded an increase in the intensity of the negative peak at 45 °C. The remainder of the CD spectra visually changed very little from temperature to temperature and there was no visible depreciation in the peaks. As temperature increases, CD spectra of many globular proteins tend towards increasing disorder (Figure 5.1 A, black line) as secondary and tertiary structure break down due to unfolding events (Corrêa and Ramos, 2009). It was expected that as the temperature increased the spectra of BAPKO\_0422 would become more similar to that expected of a random coil, however the overall  $\alpha$ -strand shape remained, suggesting that the protein is not globular and has a high degree of thermostability. Analysis of the samples using the CDSSTR algorithm in conjunction with reference set 7 showed a slight increase in % unordered structure and decrease in %  $\alpha$ -strand as the temperature increased, while analysis with reference set 4 only showed a decrease in %  $\alpha$ -strand. This suggests the BAPKO\_422 has high thermal stability. High thermal stability is a common feature of OmpA-type membrane spanning proteins. Intrinsic tryptophan fluorescence and CD studies demonstrated that *E. coli* OmpA maintained native secondary structure and a compact fold up to 80 °C (Sugawara *et al.*, 1996).

Analysis of the CD data provided evidence for the protein being folded and containing 36 - 43 %  $\alpha$ -strands, approximately 20 % turns and 33 - 37 % loop regions, which in conjunction with the thermostability of BAPKO\_0422 is indicative of a  $\alpha$ -barrel protein.

SEC-MALLS was used to estimate the molecular weight of BAPKO\_0422 in detergent micelles and determine the oligomeric state. The protein is putative a TM protein and has



detergent molecules bound in one of several potential forms (section 2.2.16.3, Figure 2.7). SEC-MALLS uses static light scattering and the refractive index to calculate the weight-averaged molecular mass, and therefore the oligomeric state. It is possible to calculate the mw and oligomeric state using the equation described in section 5.1.2 using a cuvette and the appropriate instruments without coupling to SEC, however SEC-coupling is preferred as any aggregates are separated ensuring a monodisperse sample and as all measurements made are the differences between buffer and buffer with protein sample, a continuous flow guarantees that the buffer composition is exactly the same for both the reference and experimental samples, allowing accurate determination of the baselines for all detectors (Slotboom *et al.*, 2008).

Present on the SEC-MALLS trace is a peak on the shoulder of the UV absorbance. This is indicative of smaller molecules eluting after the main peak, this is possibly due to a second species in the sample (shown to not be the case by SDS-PAGE) or smaller amounts of detergent binding to the protein. The standardised  $dn/dc$  value of 0.187 ml/g for proteins (Slotboom *et al.*, 2008) was used for the calculation of the molecular weight of BAPKO\_0422. Manual adjustment of the baselines for the light scattering detectors resulted in the molecular weight of  $22.96 \pm 9\%$  kDa, well within the error limits of the expected mw calculated using ProtParam hosted on ExPASy: 22.68 kDa (Gasteiger *et al.*, 2005). This result highlighted provided evidence that BAPKO\_0422 was monomeric and present as a monodisperse solution under the condition tested (0.01 % (w/v) LDAO).

SAXS is a low resolution technique requiring small volumes of sample. Approximately 0.15 ml BAPKO\_0422 concentrated to approximately 10 mg/ml was used to provide further characterisation of BAPKO\_0422 by calculating the radius of gyration and a molecular envelope showing the size and shape of the electron density. Whilst capable of providing a valuable insight into the overall shape of the molecule, one limitation is that there are multiple possible solutions to the molecular envelope calculation that fit the data equally well. This is compensated for by the comparison and averaging of numerous the solutions. The data can be plotted as a Guinier plot, which can be used as a good early indicator of the quality of the data; an upturn in the data at low  $q^2$  indicates protein aggregation, whilst a downturn points to interparticle interference (Jacques and Trewhella, 2009). The  $R_g$  can be calculated using the slope of the line as an early estimation of the minimum dimension of the

protein. A more accurate  $R_g$  is calculated from the  $P(r)$  plot; obtained through Fourier transform of the data.

SAXS data collection times vary depending on the strength of scattering produced by the sample. Initial test data acquisitions were done in triplicate to determine if X-ray damage had any effect on the quality of data obtained. A comparison of 3 datasets acquired sequentially on the same sample revealed no detectable change in scattering, suggesting that radiation damage would not be significant over the experimental time period.

To determine the effect of LDAO concentration on scattering, the scattering of 10 mg/ml lysozyme in two buffers (0.3 M NaCl, 30 mM Tris, pH 8.0) containing 0.1 % and 0.01 % (w/v) LDAO was measured. The concentration of LDAO was shown to increase the overall scattering (Figure 5.14) at both concentrations, this was compensated for in both cases by subtraction of a buffer-detergent blank, as evidenced by the  $R_g$  calculations in table 20. Therefore buffer containing 0.1 % (w/v) LDAO was used for acquisition of BAPKO\_0422 scattering as this was the buffer in which the protein resided after SEC. The  $R_g$  for lysozyme was calculated to be 15.6 Å from the Guinier plot and 16.39 Å from the  $P(r)$  using GNOM; the latter of which is in close agreement with that in the literature (Arai and Hirai, 1999), confirming the statement from (Jacques and Trewhella, 2009) that the  $R_g$  from Guinier analysis should be taken as a preliminary calculation and should be more accurately determined from the  $P(r)$ . Calculation of the  $P(r)$  was conducted using three different values for  $D_{max}$  to determine the accuracy of the initial value, 50 Å, chosen for the maximum width of the protein molecule. The values used for comparison, 40 Å and 60 Å were too small and too large respectively, indicating that a good choice had been made. The  $P(r)$  plot provides important information about the oligomeric state and quality of the protein sample. A plot with a  $D_{max}$  of 50 Å produced a smooth bell-shaped curve, indicating that there was no sample aggregation or interparticle interference (Jacques and Trewhella, 2009), and that at a distance of 50 Å there was no particle overlap (Roe, 2000).

From the GNOM output ten molecular envelopes were generated (Figure 5.18), evaluated (table 21) and averaged (Figure 5.18) to produce a molecular envelope. Analysis of the predicted scattering plot of this averaged molecular envelope (Figure 5.19 D) showed that it was a poor fit to the data, hence the regions of low occupancy and loosely connected atoms removed, which provided a much better fit of data (Figure 5.19 E). To produce a more

accurate molecular envelope the averaged model was used to produce a fixed core of high occupancy atoms with a substantial number of connections for production of a new molecular envelope. The resulting molecular envelope was slightly larger than the lysozyme molecule and appeared to have a slightly larger cleft than the structure (Figure 5.20 A & B, displayed with lysozyme showing the solvent accessible surface). However this may be due to the structure being determined from a rigid protein crystal and not freely floating in solution; the scattering of the molecular envelope is an excellent fit to the data (Figure 5.20 C).

SAXS datasets were collected for BAPKO\_0422 at concentrations of 6 mg/ml and 10 mg/ml. Each dataset had detergent background scattering removed by the software and was normalised to 0. Of the two datasets obtained for BAPKO\_0422 the 10 mg/ml showed a exponential increase in scattering at low  $q$ , which is indicative of aggregation. This data set was run through initial analysis twice; once unaltered and once with the aggregated portion of the data removed. The 6 mg/ml dataset was used as is.

Analysis of the Guinier approximation of the  $R_g$  for each of the three datasets calculated the  $R_g$  to be 18.6 Å from the 6 mg/ml data and 21.4 Å from both the 10 mg/datasets (table 22), however the  $R_g$  increased with the increase in concentration, and the  $I(0)$  increased disproportionately to the increase in concentration, both are which are signs of aggregation. Calculation of the  $R_g$  by GNOM however, found that there was no significant increase. PRIMUS, which did the automatic calculation noted that the full 10 mg/ml dataset was “aggregated” (table 22). Calculation of the  $R_g$  from the  $P(r)$  plot using GNOM showed very similar  $R_g$ s for the 6 and 10 (aggregation removed) mg/ml datasets, although with a difference between the two larger than the calculated errors. The whole 10 mg/ml dataset had an  $R_g$  and  $I(0)$  much larger than the other two, hence was discarded. GNOM automatically evaluates the solution calculated, finding the 6 mg/ml to be a “reasonable solution” and the 10 mg/ml to be a “good solution”. The  $I(0)$  increased with increasing concentration, but slightly more than proportionally. However as the scattering from very large protein aggregates was omitted from the calculation and hence did not contribute to the calculated solution, the data was good.

The  $P(r)$  function of BAPKO\_0422 at both concentrations was plotted with various  $D_{\max}$  values (data not shown) and the value 70 Å found to produce the best result for both datasets.

There is a 'tail' present towards high R on plots at both concentrations, this was considered to be an artefact of overlap caused by the peptide chain containing the His-tag.

The generated molecular envelopes from both datasets (Figure 5.25 & Figure 5.28) all had scores below 0.0074. The lower the score, the better; 0.0070 was the lowest score obtained by a solution from BAPKO\_0422 data. Molecular envelopes generated from the lysozyme data had similarly low scores. These scores suggested an excellent fit to the data, although comparison of these solutions to the average model (Table 5.4) resulted in solution J (Figure 5.28) being discarded due to the region of density protruding from the side. Each set of solutions was averaged and refined.

The molecular envelope produced from averaging the BAPKO\_0422 6 mg/ml dataset (Figure 5.26 A) was analysed for fit to the data (using a predicted scattering curve) by CRY SOL and found to be a very poor fit (Figure 5.26 D). This was refined using DAMFILT as with the lysozyme data and this produced a solution that had a very good fit to the experimental data, apart from at high q (labelled in the figure as s) (Figure 5.26 E). In addition there was a small region of density present at what was assigned the bottom of the solution. The solution was globular with a single protrusion at the designated "bottom". The molecular envelope of a  $\alpha$ -barrel might be expected to resemble a cylinder, however the refined electron density map calculated by DAMFILT revealed a globular structure with a single protrusion. This protrusion could be due to the extended N-terminus with a hexa-His-tag and the HRV-3C recognition sequence. This may be confirmed by recollecting SAXS data after His-tag cleavage by HRV-3C protease. The globular structure (as opposed to a cylinder) may be due to associated detergent molecules producing increased electron density surrounding the barrel. The crystal structures of *E. coli* OmpA and OmpW revealed several tightly associated detergent molecules per molecule (PDB accession codes 1BXW & 2F1T).

A fixed core of density was prepared from the averaged model and a further molecular envelope developed. This resulted in the solution showed in Figure 5.27. The molecular envelope is smaller and more compact, missing a region of density at the bottom, preventing the OmpA structure from fitting within. The 'squashed' barrel of OmpX however, would fit. The density towards the top of the model fits in line with having longer loop regions on one side and shorter on the other, however the region of density at the bottom is still there. This solution is an excellent fit compared to the experimental data (Figure 5.27 C).

The averaged molecular envelope from the 10 mg/ml dataset (Figure 5.29 A) had a predicted scattering curve that was very different to that of the recorded data (Figure 5.29 D), providing more support for the fact that simply averaging generated solutions was not accurate. As with the 6 mg/ml dataset the refined solution produced by DAMFILT (Figure 5.29 B and C) fitted the data much better (Figure 5.29 E). This molecular envelope contained more density at the bottom of the model than that of the refined 6 mg/ml solution. The structure of OmpA does not account for this, however BAPKO\_0422 is predicted to have longer periplasmic turns than OmpA, hence this area may be a result. In addition there is much less unoccupied density at the top and at the sides of the structure, suggesting a tighter fit to *E. coli* OmpA.

The solution generated by modelling based on a fixed core (Figure 5.30 A & B) has a curve very similar to that of the experimental data (Figure 5.30 C) and has very few regions of electron density that are not filled. This is not as ‘squashed’ as the molecular envelope produced from the 6 mg/ml data and is more similar to OmpA or OmpW, not OmpX. The fit of OmpA within the molecular envelope is much tighter, although there is still given for the loop regions to be flexible.

Most parameters were left at the default values, however some had to be specified, for instance  $D_{\max}$  had to be specified in production of  $P(r)$  plots. In generating the initial molecular envelopes adjustment of parameters such as the symmetry and anisotropy affected the results. The symmetry was set to the default value of P2 (no symmetry), changing this to P2 with unknown anisotropy, however resulted in a molecular envelope shaped like an elongated hexagon with a tail either side. Changing anisotropy to prolate with P2 symmetry resulted in a solution with a similar overall shape to that of a two-armed spiral galaxy. These were both deemed good fits to the data despite being unrealistic, and served to impress the importance of selecting the correct parameters for the protein. Even when the parameters are matched well, it is still possible that unrealistic solutions can be generated (as with solution J, Figure 5.28), hence it is important to generate multiple solutions and average them. The averaged solution produced by DAMAVER has been shown by CRY SOL in analysis of all datasets in this research to not be accurate to the experimental data, whereas the solution produced by DAMFILT is much more accurate.

The molecular envelope of BAPKO\_422 generated by SAXS is comparable to other 8-stranded membrane spanning proteins of known structure, in particular OmpX. The data

suggests the existence of a protrusion at the periplasmic end of the barrel which is consistent with the N-terminal tag of 37 amino acids. The larger globular domain may represent the region exposed on the cell exterior where longer loops are predicted (Figure 3.13).

These loop regions may interact with factor H and possibly other host factors e.g. factor I or factor D. Given the analogous relationship with the OmpW, OmpX and the multifunctional OmpA, it is highly likely that these proteins have multiple roles within the Spirochaete, which may include maintaining cell shape, transport of ions or hydrophobic molecules and ligand binding. It is possible that the proteins homologous to BAPKO\_0422 fulfil some or all of these roles, and / or are responsible for factor H binding in different hosts.

To increase the chances of successfully crystallising BAPKO\_0422, different screens and methods, each with its own advantages were used. The CSS screens are standard screens designed to produce protein crystals whilst having a significantly lower chance of producing salt crystals. The Memb-PASS screen (Jena Bioscience) was used as it was specifically designed to screen a wide range of membrane protein crystallisation conditions, the majority of which contained a large concentration of PEG, this however is in contrast to the conditions previously found to produce crystals of BAPKO\_0422 (section 5.2.1). Lipid cubic phase crystallography was chosen as the lipid mesophase has been reported as being an ideal environment for the growth of membrane protein crystals, this is however difficult to set up, made even more so by the fact that the laboratory lacked the recommended apparatus.

## Chapter 6. Conclusions & Future Directions

### 6.1 Conclusions

The aim of this research was to identify and characterise a family of transmembrane  $\beta$ -barrel proteins in *Borrelia* that had implications for increasing the knowledge of the Spirochaete's virulence and immune evasion mechanisms.

Several factors suggest that *Borrelia*, like Gram-negative bacteria such as *E. coli* may contain small, OmpA-like  $\beta$ -barrel TM proteins. Both have in common: a double membrane structure, a peptidoglycan layer (of differing thicknesses and positions within the periplasmic space) and the numerous lipoproteins. More substantial evidence is provided by the presence of: an orthologue to a component of the  $\beta$ -barrel assembly apparatus (BamA, BB0795) (Lenhart and Atkins, 2009) and orthologues of SecA CM secretion apparatus (Guina *et al.*, 1998). BamA is one of a complex of proteins that includes BamB through BamE (Lenhart and Atkins, 2009) present in the outer membrane of *E. coli* and *N. meningitidis* that comprises the  $\beta$ -barrel assembly machine (BAM) responsible for the folding and orientation of OM proteins. A BamA orthologue (BB0795) was recently identified in *B. burgdorferi* s.s., providing evidence for the presence of machinery responsible for the assembly of OM  $\beta$ -barrel proteins in *Borrelia* (Lenhart and Atkins, 2009). The membrane-spanning domain in this BamA orthologue shares 18.5 % identity with BamA from *E. coli*. SecA in *E. coli* is a protein responsible for export of OM proteins through the CM in *E. coli* (Guina *et al.*, 1998). With the similarities in ultrastructure, and the presence of these OM protein assembly and folding mechanisms it is logical to assume that *Borrelia* may contain small OmpA-like TM  $\beta$ -barrel proteins. A large number of non-essential *Borrelial* surface exposed proteins are encoded on plasmid genes (Steere, 2001), such as virulence factors (i.e. BbCRASPs). Proteins essential for bacterial function and survival are encoded on the chromosome, (i.e. OmpA from *E. coli*). Putative OmpA-like proteins identified in *Borrelia* are encoded on the chromosome.

The bioinformatic analysis conducted within chapter 3 of this research provides evidence for the presence of four small membrane spanning  $\beta$ -barrel proteins in each of the species of *Borrelia* analysed. A total of 12 protein homologues were predicted to be putative TM  $\beta$ -barrel proteins similar to the OmpA protein family. As was evidenced by the results from FFAS03 searching, the *Borrelial* OmpA-like proteins have very low sequence similarity (10 -

15 %) to the currently known Omps from other Gram-negative bacteria. This level of similarity is indicative of remote homology, which may be expected given the remote relationship between *Borrelia* and *E. coli*; an example of this is the orthologue of *E. coli* BamA in *Borrelia*. PRED-TMBB predicted that the majority of putative OmpA-like proteins identified would contain eight TM  $\alpha$ -strands, SignalP predicted that the majority of proteins would have a signal peptide and FFAS03 found similarities between OmpA/W and the identified proteins. Homology modelling using OmpA and OmpW from *E. coli* as templates showed that with careful alignment the sequence for BAPKO\_0422 formed  $\alpha$ -barrel proteins with an aromatic girdle.

BAPKO\_0422 was shown to exist as a monomer in 0.1 % (w/v) LDAO by size exclusion chromatography, SEC-MALLS and SAXS. The monomeric state in detergent gives further credence to the possibility that it is an OmpA-like protein (OmpA is known to readily refold in detergent).

Secondary structure analysis of BAPKO\_0422 by circular dichroism showed that the protein was approximately 35 - 40 %  $\beta$ -sheet (Table 5.1 & Table 5.2) which is comparable to known TM  $\alpha$ -barrel proteins (OmpA/X/W). Data acquisition was however limited to 195 nm. Data acquired at a lower wavelength might yield a more accurate calculation as the reference datasets and algorithm were optimised for data up to and including 190 nm.

Generation of a molecular envelope from SAXS data concurs with SEC-MALLS showing that the protein is present in a monomeric state. The molecular envelope generated fits the solvent accessible surface of *E. coli* OmpA and OmpW with only the loop regions protruding outwards. As the loop regions are flexible the molecular envelope accounts for an average of positions. The techniques and software used to produce this molecular envelope were validated by the use of lysozyme as a positive control.

Crystallisation of BAPKO\_0422 was attempted with mixed success. Crystals were obtained from lipid cubic phase crystallography, but were not of diffraction quality. Optimisation of these conditions again met with varying success, with dehydration of the drop producing clusters of unordered crystals combined with a majority of clear protein drops suggesting that a higher protein concentration was the way forward. Unfortunately due to the difficulties in obtaining pure protein, there was not enough time to attempt this. However the growth of 2D



crystals shows that it is possible to grow crystals of native BAPKO\_0422 without the need for mutation to increase contact surfaces as in the crystallisation of *E. coli* OmpA and OmpX (Pautsch *et al.*, 1999).

It is highly likely that the putative *Borrelial* OmpA-like proteins have multiple functions within the cell such as maintaining cell structure and/or functioning as a pore. A study in 2009 identified that one of the putative OmpA-like proteins identified in this research, BG0407 had the capacity to bind human complement regulating protein fH (Bhide *et al.*, 2009), although humans are an accidental host for *Borrelia*. The Spirochaete did not evolve for infectivity of humans as ticks in their natural habitat rarely encounter humans. The fH binding proteins (i.e. CRASPs) are more likely to have evolved for binding fH in birds and small mammals, hence it is possible that different groups of fH binding proteins (CRASPs, Erps, Omps) evolved to bind fH in different species.

The Bhide (2009) study speculated that BG0407 was a protein with two N-terminal coiled-coil motifs, however the data acquired throughout this research provides evidence for BAPKO\_0422 being a small TM -barrel protein. The bioinformatic studies and topology predictions predict that the protein will have a signal peptide and will form an 8-stranded TM -barrel. Each of the twelve homologues identified are currently classed as putative uncharacterised proteins (ExPASy). The fact that the protein is soluble in LDAO and monomeric as shown by SEC-MALLS is more evidence for BAPKO\_0422 being a membrane protein. The SAXS generated molecular envelope is of the right size and shape to account for a barrel-shaped protein with flexible regions. Each piece of the puzzle fits with the conclusion that BAPKO\_0422 (and the family identified within this research) are ~ 20 kDa TM -barrel proteins. Due to the differences between BAPKO\_0422 and *E. coli* OmpA/W, namely the lack of a periplasmic N-terminal domain the name proposed for this family of proteins is OmpB.

## 6.2 Future directions

The discovery of this novel family of small eight stranded membrane spanning  $\beta$ -barrel proteins is significant as this is the first time proteins belonging to the OmpA-like superfamily have been identified in the Spirochaete Phylum. *Leptospira* and *Treponema* have a related double membrane structure and a conserved  $\beta$ -barrel assembly machinery (BamA) orthologue therefore it is possible that these bacteria also contain proteins belonging to the OmpA-like superfamily. Future work could be directed towards using HMM searches in an attempt to identify these proteins in other Spirochaetes. As high-resolution structural data of *Borrelial* OmpA-like proteins becomes available this will enable improvement of the HMM profile by inclusion of these proteins.

To facilitate this, it would be beneficial to produce vector-constructs of the remaining putative OmpA-like homologues and attempt to produce high resolution structures. This would also provide an insight into the potential function of these proteins.

For improvement of cloning efficiency a different cloning technique such as blunt-end or TA cloning would be used, although this would introduce the possibility of the gene being inserted backwards, it would remove the need for restriction enzymes; some of which were found to be inefficient during this research. For these new constructs the His-tag could be switched to the C-terminus of the protein, which may allow more efficient folding, or different protein tags such as a SUMO-tag could be used.

*Borrelia* have numerous exposed proteins already known to bind fH (such as CRASPs, Erps); a vital defence mechanism for invading pathogens, as fH binding acts as an 'off-switch' for AP activation of the complement immune system. Evidence suggests that the *Borrelial* OmpA-like proteins identified in this study may bind fH from humans and/or various other hosts. It remains a possibility that the interactions between fH and BG0407 identified by Bhide *et al* (2009) may be a non-specific interaction. Isothermal titration calorimetry could be used in conjunction with ALBI assays to evaluate the potential interactions of BAPKO\_0422 and homologues with human fH and calculate the binding affinity and specificity of any interaction. High-resolution structural data such as a cocrystal structure would provide further insight.

Lipid-cubic phase crystallography showed promising results and trials at varying protein concentrations in a range of detergents would be likely to successfully grow crystals.

## Chapter 7. References

- AI, X. & CAFFREY, M. 2000. Membrane Protein Crystallization in Lipidic Mesophases: Detergent Effects. *Biophysical Journal*, 79, 394-405.
- ALAEDINI, A. & LATOV, N. 2005. Antibodies against OspA epitopes of *Borrelia burgdorferi* cross-react with neural tissue. *Journal of Neuroimmunology*, 159, 192-195.
- ALBRECHT, R., ZETH, K., SÖDING, J., LUPAS, A. & LINKE, D. 2006. Expression, crystallization and preliminary X-ray crystallographic studies of the outer membrane protein OmpW from *Escherichia coli*. *Acta Crystallographica*, F62, 415-418.
- ALEXANDER, J. M. & COX, S. M. 1995. Lyme Disease and Pregnancy. *Infectious Diseases in Obstetrics and Gynecology*, 3, 256-261.
- ALITALO, A. 2004. *Complement Evasion by Borrelia burgdorferi Spirochetes*. University of Helsinki.
- ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402.
- ANDERSON, J. F. & MAGNARELII, L. A. 2008. Biology of Ticks. *Infectious Disease Clinics of North America*, 22, 195-215.
- ANGUITA, J., RAMAMOORTHY, N., HOVIUS, J. W. R., DAS, S., THOMAS, V., PERSINSKI, R., CONZE, D., ASKENASE, P. W., RINCÓN, M., KANTOR, F. S. & FIKRIG, E. 2002. Salp15, an *Ixodes scapularis* Salivary Protein, Inhibits CD4<sup>+</sup> T Cell Activation. *Immunity*, 16, 849-859.
- ARAI, S. & HIRAI, M. 1999. Reversibility and Hierarchy of Thermal Transition of Hen Egg-White Lysozyme Studied by Small-Angle X-Ray Scattering. *Biophysical Journal*, 76, 2192-2197.
- ARONOWITZ, R. A. 2012. The Rise and Fall of the Lyme Disease Vaccines: A Cautionary Tale for Risk Interventions in American Medicine and Public Health. *The Milbank Quarterly*, 90, 250-277.
- ARORA, A., ABILDGAARD, F., BUSHWELLER, J. H. & TAMM, L. K. 2001. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nature structural biology*, 8, 334-338.
- ARORA, A., RINEHART, D., SZABO, G. & TAMM, L. K. 2000. Refolded Outer Membrane Protein A of *Escherichia coli* Forms Ion Channels with Two Conductance States in Planar Lipid Bilayers. *The journal of Biological Chemistry*, 275, 1594-1600.
- ATHREYA, B. H. & ROSE, C. D. 1996. Lyme Disease. *Current Problems in Pediatrics*, 26, 1-20.
- AUWAERTER, P. G. 2007. Point: Antibiotic therapy is not the answer for patients with persisting symptoms attributable to lyme disease. *Clinical Infectious Diseases*, 45, 143-148.
- BABB, K., EL-HAGE, N., MILLER, J. C., CARROLL, J. A. & STEVENSON, B. 2001. Distinct regulatory pathways control expression of *Borrelia burgdorferi* infection-associated OspC and Erp surface proteins. *Infection and Immunity*, 69, 4146-4153.
- BAGOS, P. G., LIAKOPOULOS, T. D., SPYROPOULOS, I. C. & HAMODRAKAS, S. J. 2004. PRED-TMBB: a web server for predicting the topology of  $\alpha$ -barrel outer membrane proteins. *Nucleic Acids Research*, 32, Web Server issue.
- BANCHEREAU, J. & STEINMAN, R. M. 1998. Dendritic cells and the control of immunity. *Nature*, 392, 245-252.
- BANE, S. E., VELASQUEZ, J. E. & ROBINSON, A. S. 2007. Expression and purification of milligram levels of inactive G-protein coupled receptors in *E. coli*. *Protein Expression and Purification*, 52, 348-355.
- BARBOUR, A. G. 1993. Linear DNA of *Borrelia* species and antigenic variation. *Trends in Microbiology*, 1, 236-239.
- BÁRCENA-URIBARRI, I., THEIN, M., SACHER, A., BUKINIS, I., BONDE, M., BERGSTRÖM, S. & BENZ, R. 2010. P66 porins are present in both Lyme Disease and relapsing fever spirochetes: A comparison of the biophysical properties of P66 porins from six *Borrelia* species. *Biochimica et biophysica Acta*, 1197-1208.
- BECKER, M., BUNIKIS, J., LADE, B. D., DUNN, J. J. & BARBOUR, A. G. 2005. Structural Investigation of *Borrelia burgdorferi* OspB, a BactericidalFab Target. *The Journal of Biological Chemistry*, 280, 17363-17370.
- BEHER, M. G., SCHNAITMAN, C. A. & PUGSLEY, A. P. 1980. Major Heat-Modifiable Outer Membrane Protein in Gram-Negative Bacteria: Comparison with the OmpA Protein of *Escherichia coli*. *Journal of Bacteriology*, 143, 906-913.
- BEN-MENACHEM, G., KUBLER-KIELB, J., COXON, B., YERGEY, A. & SCHNEERSON, R. 2003. A newly discovered cholesteryl galactoside from *Borrelia burgdorferi*. *PNAS*, 100, 7910-7918.
- BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2002. A Rotary Motor Drives Bacterial Motion. *Biochemistry*. 5th Edition ed. New York: W. H. Freeman.
- BERGFORS, T. 2003. Seeds to crystals. *Journal of Structural Biology*, 142, 66-76.

- BERNSEL, A., VIKLUND, H. & ELOFSSON, A. 2008. Remote homology detection of integral membrane proteins using conserved sequence features. *PROTEINS: Structure, Function and Bioinformatics*, 71, 1387-1399.
- BHIDE, M. R., ESCUDERO, R., CAMAFEITA, E., GIL, H., JADO, I. & ANDA, P. 2009. Complement factor H binding by different Lyme disease and relapsing fever *Borrelia* in animals and humans. *BMC Research Notes*, 2.
- BLOW, D. 2002. *Outline of Crystallography for Biologists*, OXFORD UNIVERSITY PRESS.
- BROOKS, C. S., VUPPALA, S. R., JETT, A. M. & AKINS, D. R. 2006. Identification of *Borrelia burgdorferi* Outer Surface Proteins. *Infection and Immunity*, 74, 296-304.
- BROOKS, C. S., VUPPALA, S. R., JETT, A. M., ALITALO, A., MERI, S. & AKINS, D. R. 2005. Complement regulator-acquiring surface protein 1 imparts resistance to human serum in *Borrelia burgdorferi*. *Journal of Immunology*, 1, 3299-308.
- BROSSARD, M. & WIKEL, S. K. 2004. Tick Immunobiology. *Parasitology*, 129, S161-S176.
- BROWN, E. L., GUO, B. P., O'NEAL, P. & HÖÖK, M. 1999. Adherence of *Borrelia burgdorferi* Identification of critical lysine residues in DbpA required for decorin binding. *The Journal of Biological Chemistry*, 274, 26272-26278.
- BRYCESON, A. D. M., PARRY, E. H. O., PERINE, P. L., WARRELL, D. A., VUKOTICH, D. & LEITHEAD, C. S. 1970. Louse-borne relapsing fever. *The Quarterly Journal of Medicine*, 39, 129-170.
- BU, Z., KOIDE, S. & ENGELMAN, D. M. 1998. A solution SAXS study of *Borrelia burgdorferi* OspA, a protein containing a single-layer  $\alpha$ -sheet. *Protein Science*, 7, 2681-2683.
- BUCHANAN, S. K. 1999.  $\alpha$ -barrel proteins from bacterial outer membranes: structure, function and refolding. *Current Opinion in Structural Biology*, 9, 455-461.
- BULHELLER, B. M. & HIRST, J. D. 2009. DichroCalc - circular and linear dichroism online. *Structural Bioinformatics*, 25, 539-540.
- BUNIKIS, J., NOPPA, L. & BERGSTRÖM, S. 1995. Molecular analysis of a 66-kDa protein associated with the outer membrane of Lyme disease *Borrelia*. *FEMS Microbiology Letters*, 131, 139-145.
- BURGDORFER, W. 1984. Discovery of the Lyme Disease Spirochaete and Its Relation to Tick Vectors. *The Yale Journal of Biology and Medicine*, 57, 515-520.
- BURGDORFER, W., BARBOUR, A. G., HAYES, S. F., BENACH, J. L., GRUNWALDT, E. & DAVIS, J. P. 1982. Lyme disease-a tick borne spirochetosis? *Science*, 216, 1317-1319.
- BURGDORFER, W., BARBOUR, A. G., HAYES, S. F., PÉTER, O. & AESCHLIMANN, A. 1986. Erythema chronicum migrans - a tickborne spirochetosis. *Annuaire de la Société Helvétique des Sciences Naturelles*, 1, 69-77.
- BURKOT, T. R., PIESMAN, J. & WIRTZ, R. A. 1994. Quantitation of the *Borrelia burgdorferi* Outer Surface Protein A in *Ixodes scapularis*: Fluctuations during the Tick Life Cycle, Doubling Times and Loss while Feeding. *The Journal of Infectious Disease*, 170, 883-889.
- CAFFREY, M. 2000. A lipid's eye view of membrane protein crystallization in mesophases. *Current Opinion in Structural Biology*, 10, 486-497.
- CAFFREY, M. 2003. Membrane protein crystallization. *Journal of Structural Biology*, 142, 108-132.
- CAFFREY, M. & CHEREZOV, V. 2012. Crystallizing Membrane Proteins Using Lipidic Mesophases. *Nature Protocols*, 4, 706-731.
- CASJENS, S., PALMER, N., VUGT, R. V., HUANG, W. M., STEVENSON, B., ROSA, P., LATHIGRA, R., SUTTON, G., PETERSON, J., DODSON, R. J., HAFT, D., HICKEY, E., GWINN, M., WHITE, O. & FRASER, C. M. 2000. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Molecular Microbiology*, 35, 490-516.
- CASJENS, S. R., MONGODIN, E. F., QUI, W.-G., DUNN, J. J., LUFT, B. J., FRASER-LIGGETT, C. M. & SCHUTZER, S. E. 2011. Whole-Genome Sequences of Two *Borrelia afzelii* and Two *Borrelia garinii* Lyme Disease Agent Isolates. *Journal of Bacteriology*, 193, 6995-6996.
- CASPI, R., ALTMAN, T., DALE, J. M., DREHER, K., FULCHER, C. A., GILHAM, G., KAIPA, P., KARTHIKEYAN, A. S., KOTHARI, A., KRUMMENACKER, M., LATENDRESSE, M., MUELLER, L. A., PALEY, S., POPESCU, L., PUJAR, A., SHEARER, A. G., ZHANG, P. & KARP, P. D. 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 38, D473-479.
- CASSATT, D. R., PATEL, N. K., ULBRANDT, N. D. & HANSON, M. S. 1998. DbpA, but Not OspA, Is Expressed by *Borrelia burgdorferi* during Spirochetemia and Is a Target for Protective Antibodies. *Infection and Immunity*, 66, 5379-5387.
- CDC. 2011. *Lyme Disease* [Online]. Fort Collins. Available: <http://www.cdc.gov/lyme/> [Accessed 25/08/2011 2011].

- CHACÓN, J. L. & FERREIRA, A. J. P. 2008. Development and validation of nested-PCR for the diagnosis of clinical and subclinical infectious laryngotracheitis. *Journal of Virological Methods*, 151, 188-193.
- CHARON, N. W., GOLDSTEIN, S. F., MARKO, M., HSIEH, C., GEBHARDT, L. L., MOTALET, M. A., WOLGEMUTH, C. W., RONALD J. LIMBERGER & ROWE, N. 2009. The Flat-Ribbon Configuration of the Periplasmic Flagella of *Borrelia burgdorferi* and Its Relationship to Motility and Morphology. *Journal of Bacteriology*, 191.
- CHATTOPADHYAY, A. & LONDON, E. 1984. Determination of the CMC under exclusion of the charge of the detergent. *Analytical Biochemistry*, 139, 408-412.
- CHEN, G., SEVERO, M. S., SOHAIL, M., SAKHON, O. S., WIKEL, S. K., KOTSYFAKIS, M. & PEDRA, J. H. 2012. *Ixodes scapularis* saliva mitigates inflammatory cytokine secretion during *Anaplasma phagocytophilum* stimulation of immune cells. *Parasites & Vectors*, 5.
- CHEREZOV, V., CLOGSTON, J., PAPIZ, M. Z. & CAFFREY, M. 2006. Room to Move: Crystallizing Membrane Proteins in Swollen Lipidic Mesophases. *Journal of Molecular Biology*, 357, 1605-1618.
- CHMELAR, J., CALVO, E., PEDRA, J. H. F., FRANCISCHEITTI, I. M. B. & KOTSYFAKIS, M. 2012. Tick salivary secretion as a source of antihemostatics. *Journal of proteomics*, 75, 3842-3854.
- COLEMAN, A. S., YANG, X., KUMAR, M., ZHANG, X., PROMNARES, K., SHRODER, D., KENEDY, M. R., ANDERSON, J. F., AKINS, D. R. & PAL, U. 2008. *Borrelia burgdorferi* Complement Regulator-Acquiring Surface Protein 2 Does Not Contribute to Complement Resistance or Host Infectivity. *PLoS ONE*, 3, 1-9.
- COMPTON, L. A. & JOHNSON, W. C. 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical Biochemistry*, 155, 155-167.
- COMSTOCK, L. E., FIKRIG, E., SHOBERG, R. J., FLAVELL, R. A. & THOMAS, D. D. 1993. A monoclonal antibody to OspA inhibits association of *Borrelia burgdorferi* with human endothelial cells. *Infection and Immunity*, 61, 423-431.
- CORDES, F. S., KRAICZY, P., ROVERSI, P., SIMON, M. M., BRADE, V., JAHRAUS, O., WALLIS, R., GOODSTADT, L., PONTING, C. P., SKERKA, C., ZIPFEL, P. F., WALLICH, R. & LEA, S. M. 2006. Structure-function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of *Borrelia burgdorferi*. *International Journal of Medical Microbiology*, 296, 177-184.
- CORDES, F. S., ROVERSI, P., KRAICZY, P., SIMON, M. M., BRADE, V., JAHRAUS, O., WALLIS, R., SKERKA, C., ZIPFEL, P. F., WALLICH, R. & LEA, S. M. 2005. A novel fold for the factor H-binding protein BbCRASP-1 of *Borrelia burgdorferi*. *Nature structural & molecular biology*, 12, 276-277.
- CORRÊA, D. H. A. & RAMOS, C. H. I. 2009. The use of circular dichroism spectroscopy to study protein folding, form and function. *African Journal of Biochemistry Research*, 3, 164-173.
- CORTES, D. M. & PEROZO, E. 1997. Structural dynamics of the *Streptomyces lividans* K<sup>+</sup> channel (SKC1): oligomeric stoichiometry and stability. *Biochemistry*, 36, 10343-10352.
- COUPER, D., MARGOS, G., KURTENBACH, K. & TURTON, S. 2010. Prevalence of *Borrelia* infection in ticks from wildlife in South-West England. *Veterinary Record*, 167, 1012-1014.
- COX, D. L., AKINS, D. R., BOURELL, K. W., LAHDENNE, P. & NOGARD, M. V. 1996. Limited surface exposure of *Borrelia burgdorferi* outer surface lipoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 7973-7978.
- CRIVELLATO, E., BELTRAMI, C. A., MALLARDI, F. & RIBATTI, D. 2004. The mast cell: an active participant or an innocent bystander? *Histology and Histopathology*, 19, 259-270.
- CRIVELLATO, E., RIBATTI, D., MALLARDI, F. & BELTRAMI, C. A. 2003. The mast cell: a multifunctional effector cell. *Advances in Clinical Pathology*, 7, 13-26.
- CROWLEY, J. T., TOLEDO, A. M., LARocca, T. J., COLEMAN, J. L., LONDON, E. & BENACH, J. L. 2012. Lipid Exchange between *Borrelia burgdorferi* and Host Cells. *PLOS Pathogens*, 9, 1-17.
- CUTLER, S. J. 2009. Myths, legends and realities of relapsing fever borreliosis. *Clinical Microbiology and Infection*, 15, 395-396.
- CUTLER, S. J., MOSS, J., FUKUNAGA, M., WRIGHT, D. J. M., FEKADE, D. & WARRELL, D. 1997. *Borrelia recurrentis* Characterization and Comparison with Relapsing-Fever, Lyme-Associated, and Other *Borrelia* spp. *International Journal of Systematic Bacteriology*, 47, 958-968.
- DEFOE, G. & COBURN, J. 2001. Delineation of *Borrelia burgdorferi* p66 Sequences Required for Integrin  $\alpha_3\beta_3$  Recognition. *Infection and Immunity*, 69, 3455-3459.
- DORASAMY, S., NARAINPERSAD, N., SINGH, M. & ARIATTI, M. 2012. Novel Targeted Liposomes Deliver siRNA to Hepatocellular Carcinoma Cells *in vitro*. *Chemical Biology & Drug Design*, 80, 647-656.
- EDDY, S. 2003. *HMMER User's Guide. Biological sequence analysis using profile hidden Markov models*.
- EDDY, S. R. 1995. Multiple alignment using hidden Markov models. *ISMB-95 Proceedings*, 114-120.
- EDDY, S. R. 1996. Hidden Markov models. *Current Opinion in Structural Biology*, 6, 361-365.

- EDDY, S. R. 1998. Profile hidden Markov models. *Bioinformatics Review*, 14, 755-763.
- EDDY, S. R. 2004. What is a hidden Markov model? *Nature Biotechnology*, 22, 1315-1316.
- EDDY, S. R. 2010. *HMMER User's Guide. Biological sequence analysis using profile hidden Markov models.*
- EICKEN, C., SHARMA, V., KLABUNDE, T., LAWRENZ, M. B., HARDHAM, J. M., NORRIS, S. J. & SACCHETTINI, J. C. 2002. Crystal Structure of Lyme Disease Variable Surface Antigen VlsE of *Borrelia burgdorferi*. *The Journal of Biological Chemistry*, 277, 21691-21696.
- EICKEN, C., SHARMA, V., KLABUNDE, T., OWENS, R. T., PIKAS, D. S., HÖÖK, M. & SACCHETTINI, J. C. 2001. Crystal Structure of Lyme Disease Antigen Outer Surface Protein C from *Borrelia burgdorferi*. *The Journal of Biological Chemistry*, 276, 10010-10015.
- FALLON, B. A., LEVIN, E. S., SCHWEITZER, P. J. & HARDESTY, D. 2010. Inflammation and central nervous system Lyme disease. *Neurobiology of Disease*, 37, 534-541.
- FEDER, H. M., ABELES, M., BERNSTEIN, M., WHITAKER-WORTH, D. & GRANT-KELS, J. M. 2006. Diagnosis, treatment and prognosis of erythema migrans and Lyme arthritis. *Clinics in Dermatology*, 24, 509-520.
- FERNÁNDEZ, C., HILTY, C., BONJOUR, S., ADEISHVILI, K., PERVUSHIN, K. & WÜTHRICH, K. 2001. Solution NMR studies of the integral membrane proteins OmpX and OmpA from *Escherichia coli*. *FEBS Letters*, 504, 173-178.
- FIKRIG, E., PAL, U., CHEN, M., ANDERSON, J. F. & FLAVELL, R. A. 2004. OspB Antibody Prevents *Borrelia burgdorferi* Colonization of *Ixodes scapularis*. *Infection and Immunity*, 72, 1755-1759.
- FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J.-F., DOUGHERTY, B. A., MERRICK, J. M., MCKENNEY, K., SUTTON, G., FITZHUGH, W., FIELDS, C., GOCYNE, J. D., SCOTT, J., SHIRLEY, R., LIU, L.-I., GLODEK, A., KELLEY, J. M., WEIDMAN, J. F., PHILLIPS, C. A., SPRIGGS, T., HEDBLUM, E., COTTON, M. D., UTTERBACK, T. R., HANNA, M. C., NGUYEN, D. T., SAUDEK, D. M., BRANDON, R. C., FINE, L. D., FRITCHMAN, J. L., FUHRMANN, J. L., GEOGHAGEN, N. S. M., GNEHM, C. L., MCDONALD, L. A., SMALL, K. V., FRASER, C. M., SMITH, H. O. & VENTER, J. C. 1995. Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science*, 269, 496-498 + 507-512.
- FLODEN, A. M., WATT, J. A. & BRISSETTE, C. A. 2011. *Borrelia burgdorferi* Enolase is a Surface-Exposed Plasminogen Binding Protein. *PLoS ONE*, 6, 1-9.
- FRANKE, D. & SVERGUN, D. I. 2009. DAMMIF, a program for rapid *ab initio* shape determination in small-angle scattering. *Journal of Applied Crystallography*, 42.
- FRANZ, J. K. & KRAUSE, A. 2003. Lyme disease (Lyme borreliosis). *Best Practice & Research Clinical Rheumatology*, 17, 241-264.
- FRASER, C. M., CASJENS, S., HUANG, W., SUTTON, G. G., CLAYTON, R., LATHIGRA, R., WHITE, O., KETCHUM, K. A., DODSON, R., HICKEY, E. K., GWINN, M., DOUGHERTY, B., TOMB, J.-F., FLEISCHMANN, R. D., RICHARDSON, D., PETERSON, J., KERLAVAGE, A. R., QUACKENBUSH, J., SALZBERG, S., HANSON, M., VUGT, R. V., PALMER, N., ADAMS, M. D., GOCAYNE, J., WEIDMAN, J., UTTERBACK, T., WATTHEY, L., MCDONALD, L., ARTIACH, P., BOWMAN, C., GARLAND, S., FUJII, C., COTTON, M. D., HORST, K., ROBERTS, K., HATCH, B., SMITH, H. O. & VENTER, J. C. 1997. Genomic sequence of a lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*, 390, 680-686.
- FRASER, C. M., GOCAYNE, J. D., WHITE, O., ADAMS, M. D., CLAYTON, R. A., FLEISCHMANN, R. D., BULT, C. J., KERLAVAGE, A. R., SUTTON, G., KELLEY, J. M., FRITCHMAN, R. D., WEIDMAN, J. F., SMALL, K. V., SANDUSKY, M., FUHRMANN, J., NGUYEN, D., UTTERBACK, T. R., SAUDEK, D. M., PHILLIPS, C. A., MERRICK, J. M., TOMB, J. F., DOUGHERTY, B. A., BOTT, K. F., HU, P. C., LUCIER, T. S., PETERSON, S. N., SMITH, H. O., HUTCHISON, C. A. & VENTER, J. C. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science*, 270, 397-403.
- FRASER, C. M., NORRIS, S. J., WEINSTOCK, G. M., WHITE, O., SUTTON, G. G., DODSON, R., GWINN, M., HICKEY, E. K., CLAYTON, R., KETCHUM, K. A., SODERGREN, E., HARDHAM, J. M., MCLEOD, M. P., SALZBERG, S., PETERSON, J., KHALAK, H., RICHARDSON, D., HOWELL, J. K., CHIDAMBARAM, M., UTTERBACK, T., MCDONALD, L., ARTIACH, P., BOWMAN, C., COTTON, M. D., FUJII, C., GARLAND, S., HATCH, B., HORST, K., ROBERTS, K., SANDUSKY, M., WEIDMAN, J., SMITH, H. O. & VENTER, J. C. 1998. Complete Genome Sequence of *Treponema pallidum*, the Syphilis Spirochete. *Science*, 281, 375-388.
- FREUDL, R., SCHWARZ, H., STIERHOF, Y.-D., GAMON, K., HINDENNACH, I. & HENNING, U. 1986. An Outer Membrane Protein (OmpA) of *Escherichia coli* K-12 Undergoes a Conformational Change during Export. *The Journal of Biological Chemistry*, 261, 11355-11361.

- FUJITA, T. 2002. EVOLUTION OF THE LECTIN-COMPLEMENT PATHWAY AND ITS ROLE IN INNATE IMMUNITY. *Nature Reviews: Immunology*, 346, 346-353.
- GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. R., APPEL, R. D. & BAIROCH, A. 2005. *Protein Identification and Analysis Tools on the ExPASy Server*, Totowa, NJ, Humana Press Inc.
- GEOGHEGAN, J. A., CORRIGAN, R. M., GRUSZKA, D. T., SPEZIALE, P., O'GARA, J. P., POTTS, J. R. & FOSTER, T. J. 2010. Role of Surface Protein SasG in Biofilm Formation by *Staphylococcus aureus*. *Journal of Bacteriology*, 192, 5663-5673.
- GILLET, L., SCHROEDER, H., MAST, J., THIRION, M., RENAULD, J.-C., DEWALS, B. & VANDERPLASSCHEN, A. 2009. Anchoring tick salivary anti-complement proteins IRAC I and IRAC II to membrane increases their immunogenicity. *Vetinary Research*, 40.
- GLATTER, O. & KRATKY, O. 1982. *Small Angle X-Ray Scattering*, London, ACADEMIC PRESS INC. (LONDON) LTD.
- GOUJON, M., MCWILLIAM, H., LI, W., VALENTIN, F., SQUIZZATO, S., PAERN, J. & LOPEZ, R. 2010. A new bioninformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, W695-9.
- GRANKVIST, O., WALTHER, L., BREDBERG-RÅDÉN, U., LYAMUYA, E., MHALU, F., GUSTAFSSON, Å., BIBERFELD, G. & WADELL, G. 1996. Nested PCR assays with novel primers yield greater sensitivity to Tanzanian HIV-1 samples than a commercial PCR detection kit. *Journal of Virological Methods*, 62.
- GRAY, J. S., DAUTEL, H., ESTRADA-PENÑA, A., KAHL, O. & LINDGREN, E. 2009. Effects of Climate Change on Ticks and Tick-Borne Diseases in Europe. *Interdisciplinary Perspectives on Infectious Diseases*, 2009.
- GREENFIELD, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1, 2876-2890.
- GUINA, T., HELFET-HILLIKER, D., RAMAMURTHY, V. & OLIVER, D. 1998. Sequence and phylogenetic analysis of the *Borrelia burgdorferi* secA gene. *Biochimica et Biophysica Acta* 1371, 24-30.
- GUO, B. P., BROWN, E. L., DORWARD, D. W., ROSENBERG, L. C. & HÖÖK, M. 1998. Decorin-binding adhesions from *Borrelia burgdorferi*. *Molecular Microbiology*, 30, 7111-723.
- GUO, X., BOOTH, C. J., PALEY, M. A., WANG, X., DEPONTE, K., FIKRIG, E., NARASIMHAN, S. & MONTGOMERY, R. R. 2009. Inhibition of Neutrophil Function by Two Tick Salivary Protiens. *Infection and Immunity*, 77, 2320-2329.
- HANNIER, S., LIVERSIDGE, J., STERNBERG, J. M. & BOWMAN, A. S. 2003. *Ixodes ricinus* tick salivary gland extract inhibits IL-10 secretion and CD69 expression by mitogen-stimulated murine splenocytes and induces hyporesponsiveness in B lymphocytes. *Parasite Immunology*, 25, 27-37.
- HAUPT, K., REUTER, M., ELSER, J. V. D., BURMAN, J., HÄLBICH, S., RICHTER, J., SKERKA, C. & ZIPFEL, P. F. 2008. The *Staphylococcus aureus* Protein Sbi Acts as a Complement Inhibitor and Forms a Tripartite Complex with Host Complement Factor H and C3b. *PLoS Pathogens*, 4, 1-14.
- HAYES, S. F., BURGDOFFER, W. & BARBOUR, A. G. 1983. Bacteriophage in the *Ixodes dammini* Spirochete, Etiological Agent of Lyme Disease. *Journal of Bacteriology*, 154, 1436-1439.
- HEINZE, D. M., CARMICAL, J. R., ARONSON, J. F. & THANGAMANI, S. 2012. Early Immunologic Events at the Tick-Host Interface. *PLOS ONE*, 7, 1-11.
- HERBERT, A. P., UHRIN, D., LYON, M., PANGBURN, M. K. & BARLOW, P. N. 2006. Disease-associated Sequence Variations Congregate in a Polyanion Recognition Patch on Human Factor H Revealed in Three-dimensional Structure. *The journal of biological chemistry* 281, 16512-16520.
- HERRMANN, K. W. 1966. Micellar properties of some zitterionic surfactants. *Journal of Colloid and Interface Science*, 22, 352- 359.
- HILDENBRAND, P., CRAVEN, D. E., JONES, R. & NEMESKAL, P. 2009. Review Article: Lyme Neuroborreliosis: Manifestations of a Rapidly Emerging Zoonosis. *American Journal of Neuroradiology*, 30, 1079-1087.
- HJELMELAND, L. M., NEBERT, D. W. & OSBORNE, J. C. 1983. Sulfobetaine derivatives of bile acids: Nondenaturing surfactants for membrane biochemistry. *Analytical Biochemistry*, 130, 72-80.
- HONG, H., PATEL, D. R., TAMM, L. K. & BERG, B. V. D. 2006. The Outer Membrane Protein OmpW Forms an Eight-stranded -Barrel with a Hydrophobic Channel. *The journal of biological chemistry*, 281, 7568-7577.
- HOROWITZ, H. W., AGUERO-ROSENFELD, M. E., HOLMGREN, D., MCKENNA, D., SCHWARTZ, I., COX, M. E. & WORMSER, G. P. 2012. Lyme Disease and Human Granulocytic Anaplasmosis Coinfection: Impact of Case Definition on Coinfection Rates and Illness Severity. *Clinical Infectious Diseases*.
- HOVIND-HOUGEN, K. 1984. Ultrastructure of Spirochetes Isolated from *Ixodes ricinus* and *Ixodes dammini*. *The Yale Journal of Biology and Medicine*, 57, 543-584.



- HOVIS, K. M., TRAN, E., SUNDY, C. M., BUCKLES, E., MCDOWELL, J. V. & MARCONI, R. T. 2006. Selective Binding of *Borrelia burgdorferi* OspE Paralogs to Factor H and Serum Proteins from Diverse Animals: Possible Expansion of the Role of OspE in Lyme Disease Pathogenesis. *Infection and Immunity*, 74, 1967-1972.
- HOWE, C. J., BARBROOK, A. C., KOUMANDOU, V. L., NISBET, R. E. R., SYMINGTON, H. A. & WIGHTMAN, T. F. 2003. Evolution of the chloroplast genome. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 358, 99-107.
- HUANG, X., NAKAGAWA, T., TAMURA, A., LINK, K., KOIDE, A. & KOIDE, S. 2001. Formation of the Single-layer  $\alpha$ -sheet of *Borrelia burgdorferi* OspA in the Absence of the C-terminal Capping Globular Domain. *Journal of Molecular Biology*, 308, 367-375.
- HUANG, X., YANG, X., LUFT, B. J. & KIODE, S. 1998. NMR Identification of epitopes of Lyme Disease Antigen OspA to Monoclonal Antibodies. *Journal of Molecular Biology*, 281, 61-37.
- HÜBNER, A., YANG, X., NOLEN, D. M., POPOVA, T. G., CABELLO, F. C. & NORGARD, M. V. 2001. Expression of *Borrelia burgdorferi* OspC and DbpA is controlled by a RpoN-RpoS regulatory pathway. *PNAS*, 98, 12724-12729.
- HUNTE, C., JAGOW, G. V. & SCHAGGER, H. (eds.) 2003. *Membrane Protein Purification and Crystallisation: A Practical Guide*: Academic Press.
- HYDE, F. W. & JOHNSON, R. C. 1984. Genetic Relationship of Lyme Disease Spirochetes to *Borrelia*, *Treponema* and *Leptospira* spp. *Journal of Clinical Microbiology*, 20, 151-154.
- ITO, M., NAGATA, K., KATO, Y., ODA, Y., YAMAGOE, S., SUZUKI, K. & TANOKURA, M. 2003. Expression, oxidative refolding, and characterization of six-histidine-tagged recombinant human LECT2, a 16-kDa chemotacta protein with three disulphide bonds. *Protein Expression and Purification*, 27, 272-278.
- ITZHAKI, R. F., WOZNIAK, M. A., APPELT, D. M. & BALIN, B. J. 2004. Infiltration of the brain by pathogens causes Alzheimer's disease. *Neurobiology of Aging*, 25, 619-627.
- JACQUES, D. A. & TREWHELLA, J. 2009. Small-angle scattering for structural biology - Expanding the wider frontier while avoiding the pitfalls. *Protein Science*, 19, 642-657.
- JAMES, R. W. & WALLACE, B. A. 2009. An introduction to Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy. In: WALLACE, B. A. & JAMES, R. W. (eds.) *Modern Techniques for Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy*. IOS Press.
- JANES, R. W. 2008. Reference Datasets Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy of Proteins. In: WALLACE, B. A. & JANES, R. W. (eds.) *Modern Techniques in Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy*.
- JAROSZEWSKI, L., RYCHLEWSKI, L., LI, Z., LI, W. & GODZIK, A. 2005. FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Research*, 33, 284-288.
- JEANNIN, P., MAGISTRELLI, G., GOETSCH, L., HAEUW, J.-F., THIEBLEMONT, N., BONNEFOY, J.-Y. & DELNESTE, Y. 2002. Outer membrane protein A (OmpA): a new pathogen-associated molecular pattern that interacts with antigen presenting cells-impact on vaccine strategies. *Vaccine*, 20, 23-27.
- JEFFERY, C. J. 1999. Moonlighting proteins. *TRENDS in Biochemical Sciences*, 24, 8-11.
- JOHNSON, R. C., SCHMID, G. P., HYDE, F. W., STEIGERWALT, A. G. & BRENNER, D. J. 1984. *Borrelia burgdorferi* sp. nov.: Etiological Agent of Lyme Disease. *International Journal of Systematic Bacteriology*, 34, 496-497.
- JUNCADELLA, I. J., GARG, R., BATES, T. C., OLIVERA, E. R. & ANGUITA, J. 2008. The *Ixodes scapularis* Salivary Protein, Salp15, Prevents the Association of HIV-1 gp120 and CD4. *Biochemical and Biophysical Research Communications*, 367, 41-46.
- KAISER, R., LÜCKING, G. H. & DEUSCHL, G. 1997. Generalised motor neuron disease as an unusual manifestation of *Borrelia burgdorferi* infection. *Journal of Neurology, Neurosurgery & Psychiatry*, 63, 257-258.
- KALISH, R. A., LEONG, J. M. & STEERE, A. C. 1993. Association of Treatment-Resistant Chronic Lyme Arthritis with HLA-DR4 and Antibody Reactivity to OspA and OspB of *Borrelia burgdorferi*. *Infection and Immunity*, 61, 2774-2779.
- KELLY, S. M., JESS, T. J. & PRICE, N. C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta*, 1751, 119-139.
- KHALID, S., BOND, P. J., DEOL, S. S. & SANSOM, M. S. P. 2006. Modeling and simulations of a bacterial Outer Membrane Protein: OprF from *Pseudomonas aeruginosa*. *PROTEINS: Structure, Function and Bioinformatics*, 63, 6-15.
- KIM, J. H., SINGVALL, J., SCHWARZ-LINEK, U., JOHNSON, B. J. B., POTTS, J. R. & HÖÖK, M. 2004. BBK32, a Fibronectin Binding MSCRAMM from *Borrelia burgdorferi*, Contains a Disordered Region That Undergoes a Conformational Change on Ligand Binding. *The Journal of Biological Chemistry*, 279, 41706-41714.

- KIM, K., KIM, K.-P., CHOI, J., LIM, J.-A., LEE, J., HWANG, S. & RYU, S. 2010. Outer Membrane Proteins A (OmpA) and X (OmpX) Are Essential for Basolateral Invasion of *Cronobacter sakazakii*. *Applied and Environmental Microbiology*, 76, 5188-5198.
- KIM, K. S. 2001. Escherichia coli Translocation at the Blood-Brain Barrier. *Infection and Immunity*, 69, 5217-5222.
- KOEBNIK, R. 1999. Membrane assembly of the Escherichia coli outer membrane protein OmpA: Exploring sequence constraints on transmembrane  $\alpha$ -strands. *Journal of Molecular Biology*, 285, 1801-1810.
- KONAREV, P. V., VOLKOV, V. V., SOKOLOVA, A. V., KOCH, M. H. J. & SVERGUN, D. I. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36, 1277-1282.
- KOPP, A., HEBECKER, M., SVOBODOVÁ, E. & JÓZSI, M. 2012. Factor H: A Complement Regulator in Health and Disease, and a Mediator of Cellular Interactions. *Biomolecules*, 2, 46-75.
- KRAICZY, P., HANSSEN-HÜBNER, C., KITIRATSCHKY, V., BRENNER, C., BESIÉ, S., BRADE, V., SIMON, M. M., SKERKA, C., ROVERSI, P., LEA, S. M., STEVENSON, B., WALLICH, R. & ZIPFEL, P. F. 2009. Mutational analyses of the BbCRASP-1 protein of *Borrelia burgdorferi* identify residues relevant for the architecture and binding of host complement regulators FHL-1 and factor H. *International Journal of Medical Microbiology*, 299, 255-268.
- KRAICZY, P., ROSSMAN, E., BRADE, V., SIMON, M. M., SKERKA, C., ZIPFEL, P. F. & WALLICH, R. 2006. Binding of human complement regulators FHL-1 and factor H to CRASP-1 orthologs of *Borrelia burgdorferi*. *The Middle European Journal of Medicine*, 118, 669-676.
- KRAICZY, P., SELING, A., BRISSETTE, C. A., ROSSMANN, E., HUNFELD, K.-P., BYKOWSKI, T., BURNS, L. H., TROESE, M. J., COOLEY, A. E., MILLER, J. C., BRADE, V., WALLICH, R., CASJENS, S. & STEVENSON, B. 2008. *Borrelia burgdorferi* complement regulator-acquiring surface protein 2 (CspZ) as a serological marker of human lyme disease. *Clinical and Vaccine Immunology*, 15, 484-491.
- KRAICZY, P., SKERKA, C., BRADE, V. & ZIPFEL, P. F. 2001a. Further characterisation of complement regulator-acquiring surface proteins of *Borrelia burgdorferi*. *Infection and Immunity*, 69, 7800-7809.
- KRAICZY, P., SKERKA, C., KIRSCHFINK, M., BRADE, V. & ZIPFEL, P. F. 2001b. Immune evasion of *Borrelia burgdorferi* by acquisition of human complement regulators FHL-1/reconnectin and Factor H. *European Journal of Immunology*, 31, 1674-1684.
- KRUPKA, M., RASKA, M., BELAKOVA, J., HORYNOVA, M., NOVOTNY, R. & WEIGL, E. 2007. Biological aspects of lyme disease spirochetes: Unique bacteria of the *Borrelia burgdorferi* species group. *Biomed*, 151, 175-186.
- KRUPKA, M., ZACHOVA, K., WEIGL, E. & RASKA, M. 2011. Prevention of Lyme Disease: Promising Research or Sisyphean Task? *Archivum Immunologiae et Therapia Experimentalis*, 59, 261-275.
- KUMARAN, D., ESWARAMOORTHY, S., LUFT, B. J., KOIDE, S., DUNN, J. J., LAWSON, C. L. & SWAMINATHAN, S. 2001. Crystal structure of outer surface protein C (OspC) from the Lyme disease spirochete, *Borrelia burgdorferi*. *The EMBO Journal*, 20, 971-978.
- KURTENBACH, K., HANINCOVÁ, K., TSAO, J. I., MARGOS, G., FISH, D. & OGDEN, N. H. 2006. Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nature Reviews: Microbiology*, 4, 660-669.
- LAM, T. T., NGUYEN, T.-P., MONTGOMERY, R. R., KANTOR, F. S., FIKRIG, E. & FLAVELL, R. A. 1994. Outer surface proteins E and F of *Borrelia burgdorferi*, the agent of Lyme disease. *Infection and Immunity*, 62, 290-298.
- LAMBRIS, J. D., RICKLIN, D. & GEISBRECHT, B. V. 2008. Complement evasion by human pathogens. *Nature Reviews: Microbiology*, 6, 132-142.
- LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.
- LAROCCA, T. J., CROWLEY, J. T., CUSACK, B. J., PATHAK, P., BENACH, J., LONDON, E., GARCIA-MONCO, J. C. & BENACH, J. L. 2010. Cholesterol lipids of *Borrelia burgdorferi* form lipid rafts and are required for the bactericidal mechanism of a complement-independent antibody. *Cell Host & Microbe*, 8, 331-42.
- LEMAIRE, M., CHAMPEIL, P. & MØLLER, J. V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta*, 1508, 86-111.
- LENHART, T. R. & ATKINS, D. R. 2009. *Borrelia burgdorferi* locus BB0795 encodes a BamA orthologue required for growth and efficient localization of outer membrane proteins. *Molecular Microbiology*.
- LESCOT, M., AUDIC, S., ROBERT, C., NGUYEN, T. T., BLANC, G., CUTLER, S. J., WINCKER, P., COULOUX, A., CLAVERIE, J.-M., RAOULT, D. & DRANCOURT, M. 2008. The Genome of

- Borrelia recurrentis*, the Agent of Deadly Louse-Borne Relapsing Fever, Is a Degraded Subset of Tick-borne *Borrelia duttonii*. *PLoS Genetics*, 4, 1-11.
- LI, H., DUNN, J. J., LUFT, B. J. & LAWSON, C. L. 1997. Crystal structure of Lyme disease antigen outer surface protein A complexed with an Fab. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 3584-3589.
- LI, X., NEELAKANTA, G., LIU, X., BECK, D. S., KANTOR, F. S., FISH, D., ANDERSON, J. F. & FIKRIG, E. 2007. Role of Outer Surface Protein D in the *Borrelia burgdorferi* Life Cycle. *Infection and Immunity*, 75, 4237-4244.
- LIANG, F. T., CAIMANO, M. J., RADOLF, J. D. & FIKRIG, E. 2004. *Borrelia burgdorferi* outer surface protein (*osp*) B expression independent of *ospA*. *Microbial Pathogenesis*, 37, 35-40.
- LIPFERT, J., COLUMBUS, L., CHU, V. B., LESLEY, S. A. & DONIACH, S. 2007. Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering. *Journal of Physical Chemistry*, 111, 12427-12438.
- LIU, L., DAI, J., ZHAO, Y. O., NARASIMHAN, S., YANG, Y., ZHANG, L. & FIKRIG, E. 2012. *Ixodes scapularis* JAK-STAT pathway regulates tick antimicrobial peptides, thereby controlling the agent of human granulocytic anaplasmosis. *Journal of Infectious Diseases*, 206, 1233-1241.
- LIVENGOOD, J. A. & JR, R. D. G. 2006. Invasion of human neuronal and glial cells by an infectious strain of *Borrelia burgdorferi*. *Microbes and Infection*, 8, 2832-2840.
- LOBLEY, A., WHITMORE, L. & WALLACE, B. A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, 18, 211-212.
- LORBER, B., BISHOP, J. B. & DELUCAS, L. J. 1990. Purification of octylglucoside and determination of the CMC. *Biochimica et Biophysica Acta*, 1023, 254-265.
- MADIGAN, M. T. & MARTINKO, J. M. 2006. *Brock: Biology of Microorganisms*, Pearson Prentice Hall.
- MAGRANE, M. & CONSORTIUM, T. U. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database*.
- MAKABE, K., TERESHKO, V., GAWLAK, G., YAN, S. & KOIDE, S. 2006. Atomic-resolution crystal structure of *Borrelia burgdorferi* outer surface protein A via surface engineering. *Protein Science*, 15, 1907-1914.
- MARGOS, G., GATEWOOD, A. G., AANENSEN, D. M., HANINCOVÁ, K., TEREKHOVA, D., VOLLMER, S. A., CORNET, M., PIESMAN, J., DONAGHY, M., BORMANE, A., HURN, M. A., FEIL, E. J., FISH, D., CASJENS, S., WORMSER, G. P., SCHWARTZ, I. & KURTENBACH, K. 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *PNAS*, 105, 8730-8735.
- MARGOS, G., HOJGAARD, A., LANE, R. S., CORNET, M., FINGERLE, V., RUDENKO, N., OGDEN, N., AANENSEN, D. M., FISH, D. & PIESMAN, J. 2011. Multilocus sequence analysis of *Borrelia bissettii* strains from North America reveals a new *Borrelia* species, *Borrelia kurtenbachii*. *Ticks and Tick-borne Diseases*, 1, 151-158.
- MCCONNELL, M. J. & PACHÓN, J. 2011. Expression, Purification, and refolding of biologically active *Acinetobacter baumannii* OmpA from *Escherichia coli* inclusion bodies. *Protein Expression and Purification*, 77, 98-103.
- MCDOWELL, J. V., WOLFGANG, J., TRAN, E., METTS, M. S., HAMILTON, D. & MARCONI, R. T. 2003. Comprehensive analysis of the Factor H binding capabilities of *Borrelia* species associated with lyme disease: delineation of two distinct classes of Factor H binding proteins. *Infection and Immunity*, 71, 3597-3602.
- MIKLOSSY, J. 1993. Alzheimer's disease - a spirochetosis? *Neuroreport*, 4, 841-8.
- MIKLOSSY, J. 2008. Chronic Inflammation and Amyloidogenesis in Alzheimer's Disease - Role of Spirochetes. *Journal of Alzheimer's Disease*, 13, 381-391.
- MIKLOSSY, J. 2011. Alzheimer's disease - a neurospirochetosis. Analysis of the evidence following Koch's and Hill's criteria. *Journal of Neuroinflammation*, 8, 1-16.
- MIKLOSSY, J., KRIS, A., RADENOVIC, A., MILLER, L., FORRO, L., MARTINS, R., REISS, K., DARBINIAN, N., DAREKAR, P., MIHALY, L. & KHALILI, K. 2005. Beta-amyloid deposition and alzheimer's type changes induced by *Borrelia* spirochetes. *Neurobiology of Aging*, 27, 228-236.
- MORAN, N. A. 2002. Microbial Minimalism: Genome Reduction in Bacterial Pathogens. *Cell*, 108, 583-586.
- MOSHER, D. F. 2001. A role for fibronectin in self-repair after ischemic injury. *Nature Medicine*, 7, 290-292.
- MOTALEB, M. A., CORUM, L., BONO, J. L., ELIAS, A. F., ROSA, P., SAMUELS, D. S. & CHARON, N. W. 2000. *Borrelia burgdorferi* periplasmic flagella have both skeletal and motility functions. *PNAS*, 97, 10899-10904.
- MUNITZ, A. & LEVI-SCHAFER, F. 2004. Eosinophils: 'new' roles for 'old' cells. *Allergy*, 59, 268-275.
- MUTO, A. & OSAWA, S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *PNAS*, 84, 166-169.

- MYGLAND, Å., LJØSTAD, U., FINGERLE, V., RUPPRECHT, T., SCHMUTZHARD, E. & STEINER, I. 2010. EFNS guidelines on the diagnosis and management of European Lyme neuroborreliosis. *European Journal of Neurology*, 17, 8-16.
- NAKAMIZO, S., EGAWA, G., DOI, H., NATSUAKI, Y., MIYACHI, Y. & KABASHIMA, K. 2012. Topical Treatment with Basic Fibroblast Growth Factor Promotes Wound Healing and Barrier Recovery Induced by Skin Abrasion. *Skin Pharmacol Physiol*, 26, 22-29.
- NEELAKANTA, G., LI, X., PAL, U., LIU, X., BECK, D. S., DEPONTE, K., FISH, D., KANTOR, F. S. & FIKRIG, E. 2007. Outer Surface Protein B Is Critical for *Borrelia burgdorferi* Adherence and Survival within *Ixodes* Ticks. *PLoS Pathogens*, 3, 1-11.
- NELSON, D. L. & COX, M. M. 2008. *Lehninger Principles of Biochemistry*, W. H. Freeman.
- NEW ENGLAND BIOLABS, I. *Double Digests* [Online]. Available: [http://www.neb.com/nebecomm/tech\\_reference/restriction\\_enzymes/double\\_digests.asp](http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/double_digests.asp) [Accessed 05/02/2010 2010].
- NIELSEN, H., ENGELBRECHT, J., BRUNAK, S. & HEIJNE, G. V. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*, 10, 1-6.
- NIEMANN, H. H., SCHUBERT, W.-D. & HEINZ, D. W. 2004. Adhesions and invasions of pathogenic bacteria: a structural view. *Microbes and Infection*, 6, 101-112.
- NIGROVIC, L. E. & THOMPSON, K. M. 2007. The Lyme vaccine: a cautionary tale. *Epidemiology & Infection*, 135, 1-8.
- NOGUEIRA, S. V., SMITH, A. A., QIN, J.-H. & PAL, U. 2012. A Surface Enolase Participates in *Borrelia burgdorferi*-Plasminogen Interaction and Contributes to Pathogen Survival within Feeding Ticks. *Infection and Immunity*, 80, 82-90.
- NOLLERT, P. 2005. Review: Membrane protein crystallization in amphiphile phases: practical and theoretical considerations. *Progress in Biophysics and Molecular Biology*, 88, 339-357.
- OKEMEFUNA, A. I., GILBERT, H. E., GRIGGS, K. M., ORMSBY, R. J., GORDON, D. L. & PERKINS, S. J. 2008. The regulatory SCR-1/5 and cell surface-binding SCR-16/20 Fragments of factor H reveal partially folded-back solution structures and different self-associative properties. *Journal of Molecular Biology*, 375, 80-101.
- PACHNER, A. R. 2006. *Borrelia burgdorferi* in the Nervous System: The New "Great Imitator". *Annals of the New York Academy of Sciences*, 539, 56-64.
- PAL, U., LI, X., WANG, T., MONTGOMERY, R. R., RAMAMOORTHY, N., DESILVA, A. M., BAO, F., YANG, X., PYPAERT, M., PRADHAN, D., KANTOR, F. S., TELFORD, S., ANDERSON, J. F. & FIKRIG, E. 2004a. TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. *Cell*, 119, 457-468.
- PAL, U., SILVA, A. M. D., MONTGOMERY, R. R., FISH, D., ANGUITA, J., ANDERSON, J. F., LOBET, Y. & FIKRIG, E. 2000. Attachment of *Borrelia burgdorferi* within *Ixodes scapularis* mediated by outer surface protein A. *Journal of Clinical Investigations*, 106, 561-569.
- PAL, U., YANG, X., CHEN, M., BOCKENSTEDT, L. K., ANDERSON, J. F., FLAVELL, R. A., NORGARD, M. V. & FIKRIG, E. 2004b. OspC facilitates *Borrelia burgdorferi* invasion of *Ixodes scapularis* salivary glands. *Journal of Clinical Investigations*, 113.
- PANDIRIPALLY, V., GREGORY, E. & CUE, D. 2002. Acquisition of Regulators of Complement Activation by *Streptococcus pyogenes* Serotype M1. *Infection and Immunity*, 70, 6206-6214.
- PANGBURN, M. K. 2000. Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology*, 49, 14-157.
- PAROLA, P. & RAOULT, D. 2001. Ticks and Tickborne Bacterial Diseases in Humans: An Emerging Infectious Threat. *Ticks and Tick-borne Diseases*, 32, 897-928.
- PASTER, B. J. & DEWHIRST, F. E. 2000. Phylogenetic Foundation of Spirochetes. *Journal of Molecular Microbiology and Biotechnology*, 2, 341-344.
- PASTER, B. J., DEWHIRST, F. E., WEISBURG, W. G., TORDOFF, L. A., FRASER, G. J., HESPELL, R. B., STANTON, T. B., ZABLEN, L., MANDELCO, L. & WOESE, C. R. 1991. Phylogenetic Analysis of the Spirochetes. *Journal of Bacteriology*, 173, 6101-6109.
- PAUTSCH, A. & SCHULZ, G. E. 1998. Structure of the outer membrane protein A transmembrane domain. *Nature structural biology*, 5, 1013-1017.
- PAUTSCH, A. & SCHULZ, G. E. 2000. High-resolution Structure of the OmpA Membrane Domain. *Journal of Molecular Biology*, 298, 273-282.
- PAUTSCH, A., VOGT, J., MODEL, K., SIEBOLD, C. & SCHULZ, G. E. 1999. Strategy for Membrane Protein Crystallization Exemplified with OmpA and OmpX. *PROTEINS: Structure, Function and Genetics*, 34, 167-172.

- PAVEGLIO, S. A., ALLARD, J., MAYETTE, J., WHITTAKER, L. A., JUNCADILLA, I., ANGUITA, J. & POYNTER, M. E. 2007. The Tick Salivary Protein, Salp15, Inhibits the Development of Experimental Asthma. *The Journal of Immunology*, 178, 7064-7071.
- PERUZZI, S., GORRINI, C., PICCOLO, G., CALDERARO, A., DETTORI, G. & CHEZZI, C. 2007. Human intestinal spirochaetosis in Parma: a focus on a selected population during 2002-2005. *Acta Biomed*, 78, 128-132.
- PETERSEN, T. N., BRUNAK, S., HEIJNE, G. V. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*.
- POTTS, J. R. & CAMPBELL, I. D. 1996. Structure and Function of Fibronectin Modules. *Matrix Biology*, 15, 313-320.
- POWER, M. L., FERRARI, B. C., LITTLEFIELD-WYER, J., GORDON, D. M., SLADE, M. B. & VEAL, D. A. 2006. A naturally occurring novel allele of *Escherichia coli* Outer Membrane Protein A reduces sensitivity to bacteriophage. *Applied and Environmental Microbiology*, 72, 7930-7932.
- PRASADARAO, N. V., WASS, C. A., STINS, M. F., SHIMADA, H. & KIM, K. S. 1999. Outer Membrane Protein A-Promoted Actin Condensation of Brain Microvascular Endothelial Cells Is Required for *Escherichia coli* Invasion. *Infection and Immunity*, 67, 5775-5783.
- PRASADARAO, N. V., WASS, C. A., WEISER, J. N., STINS, M. F., HUANG, S.-H. & KIM, K. S. 1996. Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. *Infection and Immunity*, 64, 146-153.
- PROCTOR, R. A. 1987. Fibronectin: A Brief Overview of Its Structure, Function, and Physiology. *Reviews of Infectious Diseases*, 9, S317-S321.
- PROSSER, B. E., JOHNSON, S., ROVERSI, P., CLARK, S. J., TARELLI, E., SIM, R. B., DAY, A. J. & LEA, S. M. 2007. Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulphate/human complement regulator factor H SCRs 6-8. *Acta Crystallographica*, F63, 480-483.
- PUGSLEY, A. P. 1993. The Complete General Secretory Pathway in Gram-Negative Bacteria. *Microbiological Reviews*, 57, 50-108.
- PURSER, J. E., LAWRENZ, M. B., CAIMANO, M. J., HOWELL, J. K., RADOLF, J. D. & NORRIS, S. J. 2003. A plasmid-encoded nicotinamidase (PncA) is essential for infectivity of *Borrelia burgdorferi* in a mammalian host. *Molecular Microbiology*, 48, 753-764.
- RADOLF, J. D., CAIMANO, M. J., STEVENSON, B. & HU, L. T. 2012. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nature Reviews: Microbiology*, 10, 87-99.
- RAMAMOORTHY, N., NARASIMHAN, S., PAL, U., BAO, F., YANG, X. F., FISH, D., ANGUITA, J., NORGARD, M. V., KANTOR, F. S., ANDERSON, J. F., KOSKI, R. A. & FIKRIG, E. 2005. The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature*, 436, 573-577.
- RAS, N. M., LASCOLA, B., POSTIC, D., CUTLER, S. J., RODHAIN, F., BARANTON, G. & RAOULT, D. 1996. Phylogenesis of Relapsing Fever *Borrelia* spp. *International Journal of Systematic Bacteriology*, 46, 859-865.
- REUTER, M., CASWELL, C. C., LUKOMSKI, S. & ZIPFEL, P. F. 2010. Binding of the Human Complement Regulators CFHR1 and Factor H by Streptococcal Collagen-like Protein 1 (Scl1) via Their Conserved C Termini Allows Control of the Complement Cascade at Multiple Levels. *The journal of Biological Chemistry*, 285, 38473-38485.
- RHODES, G. 2000. *Crystallography Made Crystal Clear. A Guide for Users of Macromolecular Models*, Academic Press.
- RIBEIRO, J. M. C. 1987. Role of saliva in blood-feeding arthropods. *Annual Review of Entomology*, 32, 463-78.
- RIBEIRO, J. M. C., ALARCON-CHAIDEZ, F., FRANCISCETTI, I. M. B., MANS, B. J., MATHER, T. N., VALENZUELA, J. G. & WIKEL, S. K. 2006. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochemistry and Molecular Biology*, 36, 111-129.
- RIBEIRO, J. M. C., MAKOUL, G. T., LEVINE, J., ROBINSON, D. R. & SPIELMAN, A. 1985. Antihemostatic, Antiinflammatory, and Immunosuppressive Properties of the saliva of a tick, *Ixodes dammini*. *Journal of Experimental Medicine*, 161, 332-344.
- RISTOW, L. C., MILLER, H. E., PADMORE, L. J., CHETTRI, R., SALZMAN, N., CAIMANO, M. J., ROSA, P. A. & COBURN, J. 2012. The  $\alpha_3$ -integrin Ligand of *Borrelia burgdorferi* is Critical for Infection of Mice but not Ticks. *Molecular Microbiology*, Accepted Article.
- RIVIERE, G. R., RIVIERE, K. H. & SMITH, K. S. 2002. Molecular and immunological evidence of oral *Treponema* in the human brain and their association with Alzheimer's disease. *Oral Microbiology and Immunology*, 17, 113-118.

- RIZZOLI, A., HAUFFE, H. C., CARPI, G., VOURCH, G. I., NETELER, M. & ROSA, R. 2011. Lyme borreliosis in Europe. *Euro Surveill*, 16, 1-8.
- ROBERTS, W. C., MULLIKIN, B. A., LATHIGRA, R. & HANSON, M. S. 1998. Molecular Analysis of Sequence Heterogeneity among Genes Encoding Decorin Binding Proteins A and B of *Borrelia burgdorferi* Sensu Lato. *Infection and Immunity*, 66, 5275-5285.
- ROCHA, E. P. C. & DANCHIN, A. 2002. Base composition bias might result from competition for metabolic resources. *TRENDS in Genetics*, 18, 291-294.
- ROE, R.-J. 2000. *Methods of X-Ray and Neutron Scattering in Polymer Science*, Oxford University Press.
- ROSA, P. A., TILLY, K. & STEWART, P. E. 2005. The Burgeoning molecular genetics of the Lyme Disease Spirochaete. *Nature Reviews: Microbiology*, 3, 129-143.
- ROSSMAN, E., KITIRATSCHKY, V., HOFMANN, H., KRAICZY, P., SIMON, M. M. & WALLICH, R. 2006. *Borrelia burgdorferi* Complement Regulator-Acquiring Surface Protein 1 of the Lyme disease spirochetes is expressed in humans and induces antibody responses restricted to nondenatured structural determinants. *Infection and Immunity*, 74, 7024-7028.
- ROST, B. 1999. Twilight zone of protein sequence alignments. *Protein Engineering*, 12, 85-94.
- RUDENKO, N., GOLOVCHENKO, M., GRUBHOFFER, L. & JR, J. H. O. 2011. *Borrelia carolinensis* sp. nov., a novel species of the *Borrelia burgdorferi sensu lato* complex isolated from rodents and a tick from the south-eastern USA. *International Journal of Systematic and Evolutionary Microbiology*, 61, 381-383.
- RUDENKO, N., GOLOVCHENKO, M., LIN, T., GAO, L., GRUBHOFFER, L. & JR, J. H. O. 2010. Delineation of a New Species of the *Borrelia burgdorferi Sensu Lato* Complex, *Borrelia americana* sp. nov. *Journal of Clinical Microbiology*, 47, 3875-3880.
- RUPPRECHT, T. A., KOEDEL, U., FINGERLE, V. & PFISTER, H.-W. 2008. The Pathogenesis of Lyme Neuroborreliosis: From Infection to Inflammation. *Molecular Medicine*, 14, 205-212.
- SAIER, M. H. 2000. Spirochetes: Evolution, Genome Analyses and Physiology. *Journal of Molecular Microbiology and Biotechnology*, 2, 339.
- ŠALI, A. & BLUNDELL, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234, 779-815.
- SAMBIR, M., IVANOVA, L. B., BRYSKIN, A. V., GODFREY, H. P. & CABELLO, F. C. 2011. Functional analysis of *Borrelia burgdorferi* *uvrA* in DNA damage protection. *FEMS Microbiology Letters*, 317, 172-180.
- SAMBROOK, J., MACCULLUM, P. & RUSSELL, D. 2000. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.
- SAPI, E., BASTIAN, S. L., MPOY, C. M., SCOTT, S., RATTELLE, A., PABBATI, N., PORURI, A., BURUGU, D., THEOPHILUS, P. A. S., PHAM, T. V., DATAR, A., DHALIWAL, N. K., MACDONALD, A., ROSSI, M. J., SINHA, S. K. & LUECKE, D. F. 2012. Characterization of Biofilm Formation by *Borrelia burgdorferi* In Vitro. *PLoS ONE*, 7, 1-11.
- SARANTIS, H. & GRINSTEIN, S. 2012. Subversion of Phagocytosis for Pathogen Survival. *Cell Host & Microbe*, 12, 419-431.
- SARMA, J. V. & WARD, P. A. 2011. The complement system. *Cell Tissue and Research*, 343, 227.
- SCHIFFERLI, J. A. 2005. Complement: a member of the innate immune system. *Springer Seminars in Immunology*, 27, 273-275.
- SCHNEIDER, M. C., PROSSER, B. E., CAESAR, J. J. E., KUGELBERG, E., LI, S., ZHANG, Q., QUORAISHI, S., LOVETT, J. E., DEANE, J. E., SIM, R. B., ROVERSI, P., JOHNSON, S., TANG, C. M. & LEA, S. M. 2009. *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature*.
- SCHRÖDER, N. W. J., SCHOMBEL, U., HEINE, H., GÖBEL, U. B., ZÄHRINGER, U. & SCHUMANN, R. R. 2003. Acylated Cholesteryl Galactoside as a Novel Immunogenic Motif in *Borrelia burgdorferi* Sensu Stricto. *The Journal of Biological Chemistry*, 278, 33645-33563.
- SCHRÖDINGER 2010. The PyMOL Molecular Graphics System. 1.3 ed.: LLC.
- SCHROEDER, H., DAIX, V., GILLET, L., RENAULD, J.-C. & VANDERPLASSCHEN, A. 2007. The paralogous salivary anti-complement proteins IRAC I and IRAC II encoded by *Ixodes ricinus* ticks have broad and complementary inhibitory activities against the complement of different host species. *Microbes and Infection*, 9, 247-250.
- SCHULZ, G. E. 2000. -Barrel membrane proteins. *Current Opinion in Structural Biology*, 10, 443-447.
- SCHUSTER-BOECKLER, B., SCHULTZ, J. & RAHMANN, S. 2004. HMM Logos for visualisation of protein families. *BMC Bioinformatics*, 5.
- SCHUTZER, S. E., FRASER-LIGGETT, C. M., CASJENS, S. R., QIU, W.-G., DUNN, J. J., MONGODIN, E. F. & LUFT, B. J. 2011. Whole-Genome Sequences of Thirteen Isolates of *Borrelia burgdorferi*. *Journal of Bacteriology*, 193, 1018-1020.

- SCHWAN, T. G., BURGDORFER, W. & GARON, C. F. 1988. Changes in Infectivity and Plasmid Profile of the Lyme Disease Spirochete, *Borrelia burgdorferi*, as a Result of In Vitro Cultivation. *Infection and Immunity*, 56, 1831-1836.
- SCHWAN, T. G. & PIESMAN, J. 2000. Temporal Changes in Outer Surface Proteins A and C of the Lyme Disease-Associated Spirochete, *Borrelia burgdorferi*, during the Chain of Infection in Ticks and Mice. *Journal of Clinical Microbiology*, 38, 382-388.
- SCHWAN, T. G., PIESMAN, J., GOLDE, W. T., DOLAN, M. C. & ROSA, P. A. 1995. Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *Microbiology*, 92, 2909-2913.
- SCHWAN, T. G., POLICASTRO, P. F., MILLER, Z., THOMPSON, R. L., DAMROW, T. & KEIRANS, J. E. 2003. Tick-borne Relapsing Fever Caused by *Borrelia hermsii*, Montana. *Emerging Infectious Diseases*, 9, 1151-1154.
- SCHWARZ-LINEK, U., WERNER, J. M., PICKFORD, A. R., GURUSIDDAPPA, S., KIM, J. H., PILKA, E. S., BRIGGS, J. A. G., GOUGH, T. S., HÖÖK, M., CAMPBELL, I. D. & POTTS, J. R. 2003. Pathogenic bacteria attach to human fibronectin through a tandem -zipper. *Nature*, 423, 177-180.
- SEDDON, A. M., CURNOW, P. & BOOTH, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochemica et Biophysica Acta*, 1666, 105-117.
- SEEMANAPALLI, S. V., XU, Q., MCSHAN, K. & LIANG, F. T. 2010. Outer Surface Protein C Is a Dissemination-Facilitating Factor of *Borrelia burgdorferi* during Mammalian Infection. *PLoS ONE*, 5, 1-8.
- SERVANT, F., BRU, C., CARRÈRE, S., COURCELLE, E., GOUZY, J., PEYRUC, D. & KAHN, D. 2002. ProDom: Automated clustering of homologous domains. *Briefings in Bioinformatics*, 3, 246-251.
- SETUBAL, J. C., REIS, M., MATSUNAGA, J. & HAAKE, D. 2006. Lipoprotein computational prediction in spirochaetal genomes. *Microbiology*, 152, 113-121.
- SHAUGHNESSY, J., LEWIS, L. A., JARVA, H. & RAM, S. 2009. Functional Comparison of the Binding of Factor H Short Consensus Repeat 6 (SCR 6) to Factor H Binding Protein from *Neisseria meningitidis* and the Binding of Factor H SCR 18 to 20 to *Neisseria gonorrhoeae* Porin. *Infection and Immunity*, 77, 2094-2103.
- SKALLOVÁ, A., IEZZI, G., AMPENBERGER, F., KOPF, M. & KOPECKÝ, J. 2008. Tick Saliva Inhibits Dendritic Cell Migration, Maturation, and Function while Promoting Development of Th2 Responses. *The Journal of Immunology*, 180, 6186-6192.
- SKARE, J. T., MIRZABEKOV, T. A., SHANG, E. S., BLANCO, D. R., ERDJUMENT-BROMAGE, H., NUNIKIS, J., BERGSTRÖM, S., TEMPST, P., KAGAN, B. L., MILLER, J. N. & LOVETT, M. A. 1997. The Oms66 (p66) Protein Is a *Borellia burgdorferi* Porin. *Infection and Immunity*, 65, 3654-3661.
- SKOGMAN, B. H., CRONER, S., FORSBERG, P., ERNERUDH, J., LAHDENNE, P., SILLANPÄÄ, H. & SEPPÄLÄ, I. 2008. Improved laboratory diagnostics of Lyme neuroborreliosis in children by detection of antibodies to new antigens in cerebrospinal fluid. *The Pediatric Infectious Disease Journal*, 27, 605-612.
- SLÁMOVÁ, M., SKALLOVÁ, A., PÁLENÍKOVÁ, J. & KOPECKÝ, J. 2011. Effect of tick saliva on immune interactions between *Borrelia afzelii* and murine dendritic cells. *Parasite Immunology*, 33, 654-660.
- SLOTBOOM, D. J., DUURKENS, R. H., OLIEMAN, K. & ERKENS, G. B. 2008. Static light scattering to characterize membrane proteins in detergent solution. *Methods*, 46, 73-82.
- SMITH, R., O'CONNELL, S. & PALMER, S. 2000. Lyme Disease Surveillance in England and Wales, 1986-1998. *Emerging Infectious Diseases*, 6, 404-407.
- SMITH, S. G. J., MAHON, V., LAMBERT, M. A. & FAGAN, R. P. 2007. A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiology Letters*, 273, 1-11.
- SÖDING, J. 2005. Protein Homology detection by HMM-HMM comparison. *Bioinformatics*, 21, 951-960.
- SONG, T. & WAI, S. N. 2009. A novel sRNA that modulates virulence and environmental fitness of *Vibrio cholerae*. *RNA Biology*, 6, 254-258.
- SPERISEN, P. & PAGNI, M. 2005. JACOP: A simple and robust method for the automated classification of protein sequences with modular architecture. *BMC Bioinformatics*, 6.
- SREERAMA, N., VENYAMINOV, S. Y. & WOODY, R. W. 1999. Estimation of the number of -helical and -strand segments in proteins using circular dichroism spectroscopy. *Protein Science*, 8, 370-380.
- STAFFORD, K. C. 2001. Ticks. *The Connecticut Agricultural Experiment Station*, 1-2.
- STEERE, A. C. 2001. Lyme Disease. *New England Journal of Medicine*, 345, 115-125.
- STEERE, A. C., BRODERICK, T. F. & MALAWISTA, S. E. 1978. Erythema Chronicum Migrans and Lyme Arthritis: Epidemiologic Evidence for a Tick Vector. *American Journal of Epidemiology*, 108.
- STEERE, A. C., COBURN, J. & GLICKSTEIN, L. 2004. The emergence of Lyme disease. *The Journal of Clinical Investigation*, 113, 1094-1101.

- STEERE, A. C., MALAWISTA, S. E., SNYDMAN, D. R., SHOPE, R. E., ANDIMAN, W. A., ROSS, M. R. & STEELE, F. M. 1977. Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three connecticut communities. *Arthritis & Rheumatism*, 200, 7-17.
- STEERE, A. C., SIKAND, V. K., MERICE, F., PARENTI, D. L., FIKRIG, E., SCHOEN, R. T., NOWAKOWSKI, J., SCHMID, C. H., LAUKAMP, S., BUSCARINO, C. & KRAUSE, D. S. 1998. Vaccination against Lyme Disease with recombinant *Borrelia burgdorferi* outer-surface lipoprotein a with adjuvant. *The New England Journal of Medicine*, 339, 209-215.
- STEWART, P. E., BYRAM, R., GRIMM, D., TILLY, K. & ROSA, P. A. 2005. The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. *Plasmid*, 53, 1-13.
- STOORVOGEL, J., BUSSEL, M. J. A. W. M. V., TOMMASSEN, J. & KLUNDERT, J. A. M. V. D. 1991. Molecular Characterization of an *Enterobacter cloacae* Outer Membrane Protein (OmpX). *Journal of Bacteriology*, 173, 156-160.
- STROP, P. & BRUNGER, A. T. 2005. Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Science*, 14, 2207-2211.
- STÜBS, G., FINGERLE, V., WILSKE, B., GÖBEL, U. B., ZÄHRINGER, U., SCHUMANN, R. R. & SCHRÖDER, N. W. J. 2009. Acylated Cholesteryl Galactosides Are Specific Antigens of *Borrelia* Causing Lyme Disease and Frequently Induce Antibodies in Late Stages of Disease. *The journal of biological chemistry*, 284, 13326-13334.
- SUGAWARA, E., STEIERT, M., ROUHANI, S. & NIKAIDO, H. 1996. Secondary structure of the outer membrane proteins OmpA of *Escherichia coli* and OprF of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 178, 6067-6069.
- SVERGUN, D., BARBERATO, C. & KOCH, M. H. J. 1995. CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, 28, 768-773.
- SVERGUN, D. I. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25, 495-503.
- SWERDLOW, R. H. 2007. Pathogenesis of Alzheimer's disease. *Clinical Interventions in Aging*, 2, 347-359.
- TABOR, S. 2001. Expression Using the T7 RNA polymerase/Promoter System. *Current Protocols in Molecular Biology*, 11, 16.2.1-16.2.11.
- TAKAYAMA, K., ROTHENBERG, R. J. & BARBOUR, A. G. 1987. Absence of Lipopolysaccharide in the Lyme disease Spirochete *Borrelia burgdorferi*. *Infection and Immunity*, 55, 2311-2313.
- TARAGEL'OVÁ, V., KO I, J., HANINCOVÁ, K., KURTENBACH, K., DERDÁKOVÁ, M., OGDEN, N. H., LITERÁK, I., KOCIANOVÁ, E. & LABUDA, M. 2008. Blackbirds and Song Thrushes Constitute a Key Reservoir of *Borrelia garinii*, the Causative Agent of Borreliosis in Central Europe. *Applied and Environmental Microbiology*, 74, 1289-1293.
- TEMPLETON, T. J. 2004. *Borrelia* Outer Membrane Surface Proteins and Transmission Through the Tick. *The Journal of Experimental Medicine*, 199, 603-606.
- TO, W. S. & MIDWOOD, K. S. 2011. Plasma and cellular fibronectin: distinct and independent functions during tissue repair. *Fibrogenesis & Tissue Repair*, 4, 1-17.
- TOLEDO, A., COLEMAN, J. L., KUHLOW, C. J., CROWLEY, J. T. & BENACH, J. L. 2011. The Enolase of *Borrelia burgdorferi* Is a Plasminogen Receptor Released in Outer Membrane Vesicles. *Infection and Immunity*, 80, 359-368.
- TYSON, K., ELKINS, C., PATTERSON, H., FIKRIG, E. & SILVA, A. D. 2007. Biochemical and Functional Characterization of Salp20, an *Ixodes scapularis* Tick Salivary Protein that Inhibits the Complement Pathway. *Insect Molecular Biology*, 16, 469-479.
- VALENZULA, J. G., CHARLAB, R., MATHERS, T. N. & RIBEIRO, J. M. C. 2000. Purification, Cloning, and Expression of a Novel Salivary Anticomplement Protein from the Tick, *Ixodes scapularis*. *The Journal of Biological Chemistry*, 275, 18717-18723.
- VANAKEN, T., FOXALL-VANAKEN, S., CASTLEMAN, S. & FERGUSON-MILLER, S. 1986. Alkyl glycoside detergents: Synthesis and applications to the study of membrane proteins. *Methods in Enzymology*, 125, 27-35.
- VANDEPUTTE-RUTTEN, L., KRAMER, R. A., KROON, J., DEKKER, N., REGMOND, M. & GROS, P. 2001. Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *The EMBO Journal*, 20, 5033-5039.
- VATANDOOST, H., GHADERI, A., JAVIDIAN, E., NIA, A. H. Z., RASSI, Y., PIAZAK, N., KIA, E. B., SHAEGHI, M., TELMODARREIY, Z. & ABOLHASANI, M. 2003. Distribution of Soft Ticks and Their Infection with *Borrelia* in Hamadan Province, Iran. *Iranian Journal of Public Health*, 32, 22-24.
- VOGT, J. & SCHULZ, G. E. 1999. The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Structure*, 7, 1301-1309.



- VOLKOV, V. V. & SVERGUN, C. I. 2003. Uniqueness of ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, 36, 860-864.
- VUYYURU, R., LIU, H., MANSER, T. & ALUGUPALLI, K. R. 2011. Characteristics of *Borrelia hermsii* infection in human hematopoietic stem cell-engrafted mice mirror those of human relapsing fever. *PNAS*, 108.
- WANG, G., DAM, A. P. V., SCHWARTZ, I. & DANKERT, J. 1999. Molecular Typing of *Borrelia burgdorferi* Sensu Lato: Taxonomic, Epidemiological, and Clinical Implications. *Clinical Microbiology Reviews*, 12, 633-653.
- WHITMORE, L. & WALLACE, B. A. 2004. DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32, W668-673.
- WHITMORE, L. & WALLACE, B. A. 2008. Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers*, 89, 392-400.
- WIKEL, S. K. 1996. Host immunity to ticks. *Annual Review of Entomology*, 41, 1-22.
- WRIGHT, D. J. M. 2009. Borrel's accidental legacy. *Clinical Microbiology and Infection*, 15, 379-399.
- YANG, X., PROMNARES, K., QIN, J., HE, M., SHRODER, D. Y., KARIU, T., WANG, Y. & PAL, U. 2011. Characterisation of Multiprotein Complexes of the *Borrelia burgdorferi* Outer Membrane Vesicles. *Journal of Proteome research*, 10, 4556-4566.
- YANG, X. F., LYBECKER, M. C., PAL, U., ALANI, S. M., BLEVINS, J., REVEL, A. T., SAMUELS, D. S. & NORGARD, M. V. 2005. Analysis of the ospC Regulatory Element Controlled by the RpoN-RpoS Regulatory Pathway in *Borrelia burgdorferi*. *Journal of Bacteriology*, 187, 4822-4829.
- YANG, X. F., PAL, U., ALANI, S. M., FIKRIG, E. & NORGARD, M. V. 2004. Essential Role for OspA/B in the Life Cycle of the Lyme Disease Spirochete. *Journal of Experimental Medicine*, 199, 641-648.
- ZGAIR, A. K. 2012. *Escherichia coli* flagellin stimulates pro-inflammatory immune response. *World Journal of Microbiology and Biotechnology*, 28, 2139-2146.
- ZHANG, M., GAO, X., YU, Y., AO, J., QIN, J., YAO, Y. & LI, Q. 2007. Detection of Roundup Ready soy in highly processed products by triplex nested PCR. *Food Control*, 18, 1277-1281.

## Chapter 8. Appendix

### 8.1 Redundancy reduced lists of known OM proteins

A redundancy reduced list of known Outer Membrane Proteins acquired from ProDom

#### OMPs with 40% redundancy reduction

Sequences kept: 23 out of 167

Reduction: 13% of original set kept

>Q9RM69_ERWCA#29#207	Q6W821_SHISO#29#195	Q6PNM6_SHIBO#29#195
Q3Z3G4_SHISS#29#195	OMPA_ECOLI#29#195	OMPA_ECO57#29#195
Q9L6J0_ECOLI#29#195	Q49P73_ECOLI#29#195	Q49P71_ECOLI#29#195
OMPA_SHIDY#29#200	Q32HV0_SHIDY#29#199	Q8CW76_ECOL6#58#227
Q83RX2_SHIFL#29#198	Q31YM0_SHIBO#29#198	Q49P72_ECOLI#29#199
Q5PGD5_SALPA#29#199	Q57QT3_SALCH#29#199	OMPA_SALTY#29#199
OMPA_SALTI#29#199	Q52JK5_ENTSA#29#195	OMPA_ENTAE#29#197
OMPA_KLEPN#17#191	Q7N602_PHOLL#29#209	OMPA_SERMA#29#207
OMPA_YERPS#29#203	OMPA_YERPE#29#203	Q6D6D4_ERWCT#29#212

YTGKGLGVQSFDHTGFYGNQYTDVNNNPIKSKLGAGAFVGYQANPYLGFEMGYDWLGRMKYAGSTANPADSASLKAQGIQLAAKLSYPVLPDL  
VYTRLGGMVWRVDTHADRSGNHLNDDTGVSPPLAAIGIEYAIIDKNWATRVDYQWVS  
NIGDAGTVGARPDNMLMSVGLSYRFGQDD

>Q6PSK4_PASTR#28#221	Q8RJD0_HAEGA#29#211	Q8RML8_HAEGA#29#211
Q8RMM2_HAEGA#29#211	Q8RMM0_HAEGA#29#211	Q8RMM1_HAEGA#29#211
Q8RML9_HAEGA#29#211	Q5FY67_HAEGA#29#211	Q8RJ58_HAEGA#29#206
Q6TRS4_HAEGA#29#210	Q6IV12_HAEGA#29#210	Q6GUB7_PASMU#29#218
Q9CMN1_PASMU#29#220	Q6GUB4_PASMU#29#220	Q6GUB5_PASMU#29#220
Q6GUB8_PASMU#29#204	Q6GUB6_PASMU#29#215	Q8VPL3_PASMU#29#215
O51841_ACTAC#29#213	Q9S5J9_ACTAC#29#213	Q65T85_MANSM#39#236
OMP52_HAEIN#29#219	OMP51_HAEIN#29#219	Q4QLD3_HAEI8#29#225
OMP53_HAEIN#29#225	Q65T83_MANSM#29#213	Q6XAX7_PASHA#27#228
Q6XAX6_MANGL#27#228	Q6XAX8_PASHA#27#228	Q6XAX4_MANGL#27#228
Q6XAY0_PASHA#27#233	Q6XAY1_PASHA#27#233	Q6XAY2_PASHA#27#233
Q6XAX9_PASHA#27#234	Q6XAY3_PASHA#27#235	Q6XAY6_PASHA#27#239
Q9RQ28_PASHA#27#239	Q6XAY4_PASHA#27#239	Q6PSJ7_PASTR#28#225
Q6PSK1_PASTR#28#224	Q6PSK5_PASTR#28#224	Q3EF13_ACTSC#29#219

YVGAKAGWASFHDGISQIDHKNKGKYGINKNSVTYGAFFVGYQIIDNLAAEVGYEYFGRVRGLEQTKAGGSKKTFRHSAGHTTIALKGNYEVISG  
LDITYAKAGIALVNNSYKTVNVDTKVPTKTSRFQSSLILGAGVEYAITPSLGARIEY  
QWLNAGKASYATLRRMGVEGSDYRPDISSVSAGLTYRFGQGAA

>Q5DYH6\_VIBF1#21#185

YLGQGVGVSHFLGACSSNAIECKNYVTGGGLYGGYQFNSWFALEGSWHDYGSPKTFYGVGDGYYSNATGVDLSVKLSLPVTDNLDLYAKGGAA  
NYLSVSGNDNVHLGMFESSIDDIWEIGAAYALPNWSLRFGTSTIIDGIGNAKTG  
KSDLYFTSLGLTYKF

>Q8E832\_SHEON#33#191

YAGAGLGQGHYSNGSNPQSYDSVRDRFAGSVYLYGVNPNYLAPELSYQFLGSAYANYEQGQISGDFQQVVLAAARFGYPLTTSLYPYVKVGAGW  
FGDSEGLRSGSERGFSPIVAAGVEYAFPTRLSGRLEHQYTDSLGAADSIGYTDHHLT  
TLGLSWRFG

>P96773\_HAEDU#126#257

KGDLNRQFKHTAHGANLSLKPSYEILPNLDVYGVGMGLVRNDYKFILDLQSKDQHVKDETKLHTLKPSLLLTGTGVEYAITPELALRTEYQYLN  
KAGNLYKAAKYKNISTSLAYAPDIHSVSVGLSYRFGQG

>O31154\_VIBCH#24#173                      Q87RL4\_VIBPA#23#171

YVGKGVKGSWLDLACLAGQSCEDDDQVVGAFGLGYQANKWLSLEAGYDYLKGFATAAGLNDEKQVAVTLAPKLSIPLTEGIALYGVGGAYVDYGS  
KDDYSYLGAAGLEFNTNHNVTMRLEYQNLTDINNDIVRARAETATLGIAKFGGSE

>Q87JK1\_VIBPA#26#179                      Q8D6H3\_VIBVU#30#182                      Q7MDC9\_VIBVY#30#182  
Q87GX9\_VIBPA#27#181

YIGGKVGYNLNDACYLNEPCDDDSFAAGMHIGYNFNEYVAAEYGVLDYLGDFNTANFKKPGLNTVDGDLWALTLPKFNPLPLTDSWNLFQKVGAA  
YMMAGDEKDFVPTGSLGAEYKINYNWSLRAEYQRYQDMSDDIVDDMSDFFGIGFN  
YKFG

>Q44GL1\_CHRSL#62#200

DVDDDDNTFKGFVGYNFRYFATEAFYSDLRVVKLGNGTANTDLESEAYGVSLVGKLPITQWFELFAKAGMAKWEVDIDGNLGGASTDLEDND  
GVDPVYGAGAQFNFKPFLVRAEYERYDFDSYQIDSFTASVWGQF

>Q5FA01\_NEIG1#20#175                      Q9R2R1\_NEIME#20#174                      Q9RP16\_NEIME#20#174  
Q9RP18\_NEIME#20#174                      Q7AR60\_NEIMA#23#174                      P95372\_NEIME#23#174  
Q9RP17\_NEIME#23#174                      Q7DDM2\_NEIMB#23#174                      P96943\_NEIME#23#174  
P95371\_NEIME#20#175                      P95343\_NEIGO#20#174

EGASGFYVQADAHAHAKASSLSGSAKGFSPRISAGYRINDLRFAVDYTRYKNYKQAPSTDFKLYSIGASVIYDFDTQSPVKPYFGARLSLNRASA  
HLGSDSFSKTSAGLGLVLAGVSYAVTPNVDLDAGYRYNYVGKVNNVKNVRSSELGA  
GVRVKF

>Q3EI33\_ACTSC#24#191

YVEGDAGYSKLRSTGGLSSKISDGSFSPSAVGVYKVNDRFALDYTYYGKDEGFGDLGGSGENEVKAYGFLAAYYDFDLGTSITPYLGARIS  
ANHIKVNDSFNNASDYFHFNSNTEFGYGAIAGVSYNFAPQWDLNVAEYNRLGEV  
RDVKVNQYGAKVGVRYTF

>Q3QBC3\_9GAMM#28#162

AKGGYQWASDDSYKHSNPKGAIFGVSSGLQFSPAWSWDVGYQYHDDLKADATSVNVKAGLIESALRYDWYLDQNLVSVYGRGLGVAYWDMKTHLS  
SDKLDATGLSPLGEVGVNYNFTPNVRLSAGYQYIDSIGESN

>Q3QLJ4\_9GAMM#38#150

ASGDGEGSSDGVGFAYGGYNFNEFWGLEANLFASGDLGEDGVVDVGAGALTFTPKFTYHFNDTFSFAKVGVASMAVVVDYGPFDADYTGWGLT  
WGVGVNAALTEKLNRLSY

>Q3GMA6\_9GAMM#36#179

YVGAKVGQFDLDVDGADDPATAYGVYAGYNFDPNFGMEAEFVGSDDADYRGGDIDAKTYGAYGTYRYAFPNTGLYAKGKLGVARTEIEATGTSPA  
GFYNGSNKDTSLAGGVGLGYSVNPFSVEAEYDKLGSDADLMTVGAQLKF

>Q6D3I0\_ERWCT#17#171                      Q3Z3X1\_SHISS#19#171                      Q32I92\_SHIDY#19#171  
Q323X9\_SHIBO#19#171                      OMPX\_SHIFL#19#171                      OMPX\_ECOLI#19#171  
OMPX\_ECOL6#19#171                      OMPX\_ECO57#19#171                      Q8XH17\_SALTI#19#171  
Q7CQV8\_SALTY#19#171                      Q5PG13\_SALPA#19#171                      Q57RC6\_SALCH#19#171  
OMPX\_ENTCL#19#172                      Q48427\_KLEPN#25#170                      Q83ZR1\_ENTAE#25#171  
Q8D0S1\_YERPE#20#176                      Q8ZDQ0\_YERPE#18#174                      Q669E5\_YERPS#18#174  
Q54470\_SERMA#27#172                      Q5J664\_SERMA#27#172

GAGSAFAGQSTVTAGYAQGDAQGVQNKVKGFNLKYRYEQDNNPLGVISSFTYLEKNGSDDGFYSKGQYMGFTAGPAYRLNDWASLYGVVGFSGH  
KVTSNNTNGQDNASNDYGFYTYGAGVQFNPIQDVALDVGYEQSRIRSVDVGAWAVG  
VGYRF

>Q66AL7\_YERPS#27#179                      Q7N474\_PHOLL#26#175                      Q70AM3\_YEREN#32#178  
AIL\_YEREN#32#178                      Q7N473\_PHOLL#27#181                      Q8ZEI0\_YERPE#27#179

ASNTVSFGYAQSTLKIIDEKIGKDNKGFNLYRHELDVGLIVASFTHTKQNYGMPGSDGKRKVEYYSLMVGPSWRNFVFSAYALIGATQGG  
STHTKPRMVSNTVSKTSMGYGAGLQFNPKHVAIDTAYEYAKIEDVKIGTWIVGVG  
YRF

>Q8Z920\_SALTI#29#174                      Q9RFV4\_SALTY#29#174                      Q8ZRH3\_SALTY#29#174  
Q57SL6\_SALCH#29#174                      Q5PFY4\_SALPA#29#174

SFGYAQTHLSSLKNSDSKDLRGFNFKYRYEFNETWGMLGSFTATRNMENYTWKEGLHKNGSDSDYDGLMFGPTYRFNDYVSLYGNAGIATM  
KFNKHSKEESFAYGAGVIFNPVKISISIDASWEASRFFAVDTNTFGVSVGYRF

>Q8XBY9_ECO57#13#177	Q6DB56_ERWCT#23#182	Q9EUG9_SALEN#29#185
Q04817_SALTY#29#185	Q5PMG8_SALPA#31#185	Q8Z6B2_SALTI#31#185
Q57Q49_SALCH#31#185	PAGC_SALTY#31#185	Q6WPL6_ECOLI#13#177
Q7AAV5_ECO57#18#182		

SAGYAQSKVQDFKNIKGVNLQYRYEWDSPVSVVGSFSYMKGDWADSHRDEADDFYRHQADIKYYSFLAGPAYRLNDYISFYGLVGISHTKAKGD  
YEWNRNSVGADES DGYLSESVSKKSTDFAYAAGVI INPWGNMSVNVGYEGTKADIYG  
KHSVNGFTVGVGYRF

>Q9XD46\_VITS1#32#191

TVSIGYSHGKISGADKLNVTAKYNYQFDQQPWGVMNTLYMGGKQRNNITSSNQLYENDVDVDYYSAGVGPSYRINPNVNVYGVGVAKADTK  
GTQKQNRGTGRIYNIIDNIDTNVMYGAGVQYNPAPNWSVDVGYESSRVNDGYDKRSMN  
AFNVGVGYRF

>Q8X428_ECO57#7#108	Q8FIF5_ECOL6#32#199	Q3Z1B4_SHISS#32#199
Q8FEX5_ECOL6#32#199	Q322C6_SHIBO#19#199	Q81ZK9_SHIFL#19#199
Q320M5_SHIBO#19#199	Q8X547_ECO57#19#199	Q8X282_ECO57#19#199
Q8X2A1_ECO57#19#199	Q8XEG5_ECO57#19#199	Q7AFX4_ECO57#19#199
Q687E7_9VIRU#19#199	Q6H9T1_9CAUD#19#199	Q8X281_ECO57#29#199
Q9EV15_ECO57#29#199	Q7DBL7_ECO57#29#199	

SVMAGPSVRVNEWFSAYAMAGVAYSRVSTFXGDYLRVTDNKGKTHDVLTGSDDGHSNTSLAWGAGVQFNPXESVAIDIA YEGSGSGDWRTDGF  
IVGVGYKF

>O33796_SALTY#23#153	Q8ZQ85_SALTY#23#153	Q57QV9_SALCH#23#153
----------------------	---------------------	---------------------

AGYKNTVSI GYAYTDLSGWLSGNANGANIKYNWEDLDSGFGAMGSVTYTSADVNNYGYKVGADADYTSLLVGPSYRFNDYLNAYVMIGAANGHIK  
DNWGNSDNKTA FAYGAGIQLNPVENIAVNASYEHTSF

>Q7N5I2\_PHOLL#43#182

LNGATLSYRYELDDQWGLSSFTFAKGDEKETKYTGEDLKYYSSVMIGPTYRINDYISLYGQLGLSRINDKSTAHYSGGYIEKESTSKNTLWG  
AGFIINPTNTSITAGYEGSRFSIKDGNKDHLS TNGFNITVGYRF

>Q57K85_SALCH#72#178	Q8ZM87_SALTY#69#178	Q8Z3Y8_SALTI#72#178
GSQRYSDSSNGRVTTTRYSLLAGPSWKINNQLSLYSQVGPVLLHQRDHGINESDSKVGYSAGVAYTPVSNVAITLGYEGADFDATHNSGSL NSNGFN LGVGYRF		

>Q8ZQG1_SALTY#92#172	Q8D0Z7_YERPE#103#194	Q8ZCR3_YERPE#91#182
Q74SP4_YERPE#91#182	AIL_YERPS#91#182	Q8ZF58_YERPE#98#183
Q66BP0_YERPS#98#183		

GPSYRFNEYLNAYVMAGLGHGHIDDKRDN SGKKTGFAYGAGVQINPVENIAINASYEYSRFSAYDSKVNAGTWVLGVGYSF

The second phase of redundancy reduction. After the initial reduction, each half of the list was recombined and resubmitted. No more sequences were removed.

### OMPs with 90 % redundancy reduction

Sequences kept: 68 out of 167

Reduction: 41% of original set kept

>OMPA\_SHIDY#29#200  
YTGAKLGWSQYHDTGFI DNNGP THENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQLTAKLGYPITDDLDVYTRLGGMV  
WRADTKAHNNVTGESEKNHDTGVSPVFAGGV EWAITPEIATRLEYQWTNNIGDAHT  
IGTRPDNGLLSLGVSYRFQQGE

>OMPA\_SALTI#29#199  
YAGAKLGWSQYHDTGFIHNDGP THENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGDNTNGAYKAQGVQLTAKLGYPITDDLDVYTRLGGMV  
WRADTKSNVPGGASTKDHD TGVS P VFAGGIEYAITPEIATRLEYQWTNNIGDANTI  
GTRPDNGLLSVGVSYRFQQGE

>Q52JK5\_ENTSA#29#195  
YAGGKLGWSQFHDTGFI PNDGP THESQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGDVTNGAFKAQGVQLTAKLGYPVTDDLDVYTRLGGMV  
WRADSSSNIAGDDHDTGVSPVFAGGV EAMTRDIATRLEYQWVNNIGDAQTVGARP

DNGMLSVGVSYRFGQQE

>OMPA\_ENTAE#29#197  
YAGGKLGWSQFHDGTGFWYNSLNNNGPHTHESQLGAGAFGGYQVNPYLGFMGYDWLGRMPYKGVKVNGAFSSQAVQLTAKLGYPITDDLDIYTRL  
GGMVWRADSSNSIAGDNHDTGVSPVFAGGVEWAMTRDIATRLEYQWVNNIGDAGTV  
GVRPDNGMLSVGVSYRFGQ

>OMPA\_KLEPN#17#191  
YAGGKLGWSQYHDTGIFYGNFGQNNNGPTRNDQLGAGAFGGYQVNPYLGFMGYDWLGRMAYKGSVDNGAFKAQGVQLTAKLGYPITDDLDIYTR  
LGGMVWRADSKGNYASTGVSRSEHDTGVSPVFAGGVEWAVTRDIATRLEYQWVNNI  
GDAGTVGTRPDNGMLSLGVSYRFGQ

>Q7N602\_PHOLL#29#209  
YTGGKLGWSQYHDTVNFYGNGYTSAIGNGPHTKNQLGAGAFGLGYQANPYLGFEGYDWLGRMPYKGSVNNGAFKAHGFQMATKLSYPIMDDLDIY  
TRLGGMVWRADSKATYSATNTHLKNHDTGVSPPLAAIGVEYAITKDWATRLDYQWVN  
NIGDAGSVGARPDNGLLSVGVSYRFGQDEAA

>OMPA\_SERMA#29#207  
YTGAJLGWSQYHDTGIFYGNFYQNGIGNPHTHKDQLGAGAFGLGYQANQYLGFEFGYDWLGRMPYKGSVNNGAFKAQGVQLAAKLSYPIADDLDIY  
TRLGGMVWRADSKANYGRGTGQRLSDHDTGVSPPLAAVGVEYALTKNWATRLDYQFVS  
NIGDAGTVGARPDNTMLSLGVSYRFGQDD

>OMPA\_YERPE#29#203  
YTGGKLGWSQYQDTGSIINNDGPTHKDQLGAGAFFGYQANQYLGFMGYDWLGRMPYKGDINNGAFKAQGVQLAAKLSYPIVQDLDVYTRLGGL  
VWRADAKGSFDDGLDRASGHDTGVSPPLVALGAEYAWTKNWATRMEYQWVNNIGDRE  
TVGARPDNGLLSVGVSYRFGQEDAA

>Q6D6D4\_ERWCT#29#212  
YTGGKLGWSQFHDGTGIFYGNGYTGVNNNPIKSKLGAGAFVGYQANPYLGFELGYDWLGRMKYAGSTASNDSSASFKAQGIQLAAKLSYPIVMPDLD  
VYTRLGGMVWRADSHAATSVNTGTRADINNDTGVSPPLAAIGVEYAIIDKNWATRVD  
YQWVSNIGDAGTVGARPDNLLMSVGLSYRFGQDD

>Q9RM69\_ERWCA#29#207  
YTGGKLGWSQFHDGTGIFYGNGYTGVNNNPIKSKLGAGAFVGYQANPYLGFEMGYDWLGRMKYAGSTANPADSASLKAQGIQLAAKLSYPIVLPDLD  
VYTRLGGMVWRVDTHADRSGNHLNDDTGVSPPLAAIGIEYAIIDKNWATRVDYQWVS  
NIGDAGTVGARPDNMLMSVGLSYRFGQDD

>Q6IV12\_HAEGA#29#210  
YAGAKAGWASFHDGLNQFENSQNAYGTLRNSVTYGVFGGYQITDNFAVELGYDDFGRAKLRKGGETVIKHTNHGAHLSLKASYPVLGLDVYAR  
VGAALIRSDYKSTKRAESDYVMHEHSLKVSPVFAGGLEYNLPSLPELALRVEYQWV  
NKVGRGEKDGSRVDYTPSIGSVTAGLSYRFGQ

>Q6GUB7\_PASMU#29#218  
YVGAKAGWASFHDGLNQFNDITEPATGFKRNSVTYGVFGGYQITDNFAVELGYDDFGRAKARATDPKTKETVDAAKHTNHGAHLSLKASYPVL  
GLDVYARVGAALVRSDYKVYDKEPADLSFLKRTHSTQVSPVFAGGLEAYAFMPPEL  
RVEYQWLNNVGKLDKAGGERVDYRPDIGSVTAGLSYRFGQ

>Q6GUB5\_PASMU#29#220  
YVGAKAGWASFHDGLNQAKYLEAPEATFGFKRNSVTYGVFGGYQITDDFAVELGYDDFGRAKLRIAETDQKARDVAKHTNHGAHLSLKASYPVL  
DGLDVYARVGAALIRSDYKVYDHSDBAKLPQLKRTHSTQVSPVFAGGLEAYAFMPPEL  
ALRVEYQWVNNVGKLDVDGNRVDYRPDIGSVTAGLSYRFGQ

>Q6GUB8\_PASMU#29#204  
YVGAKAGWASFHDGLNQAKYLEAPEATFGFKRNSVTYGVFGEYQITDNFAVELGYDDFGRAKLRMAEKDQAKADAAKHTNHGAHLSLKASYPVL  
DGLDIYARVGAALIRSDYKVYDHSDBAKLPQFKRTHSTQVSPVFAGGLEAYAFMPPEL  
ALRVEYQWVNNVGKVKDVLGERVDYR

>Q8VPL3\_PASMU#29#215  
YVGAKAGWASFHDGLNQIEYVSKTSFGSKRNSVTYGVFGGYQITDNFAVELGYDDFGRAKVRANSKTTAFDAKHTNHGTHLSLKASYPLDGL  
DVYARVGAALIRSDYKLYAPLINKRLSPHFKITQVSPVFAGGLEAYAFIPELTLRVE  
YQWVNNVCKFEYADGQYADFRPDIGSVTAGLSYRFGQ

>Q9S5J9\_ACTAC#29#213  
YAGAKAGWASSHHGLNQFKQKGVNINRNSEAYGVFGGYQITDNFAVEAGYEFGRSKAKVNGAQRFRHTAHGTTLALKASYPVLNLDVYGRVG  
AALIRSDYKVGQANKPDYHNLKVSPVFAGGVEYAILPELALRAEYQWVSRVGNLG  
RAEEKADRSAARTIDYSPDIGSVAVGLSYRFGQGVA  
>Q65T85\_MANSM#39#236

YAGAKAGWASFHDGYTQYAEEDGVGSHTKSVTYGVFGGYQIFNRDNLGLAVELGYDDFGRALRTNGATSAKHTNHGAHLSLKPSYDLGALAPVL  
SGLDVYGVGAALVRSDYKVNDGYSYGFNKSDFADHSLKTSLLLGALEYALPSLP  
ELAFRLEYQWLNKVGKLENANGTRFDYTPETIHSVTAGVSYRFGQGVAA

>OMP52\_HAEIN#29#219  
YAGVKAGQASFHDGLRALAREKNVGYHRNSFTYGVFGGYQILNQNNLGLAVELGYDDFGRAKGREKGKTVAKHTNHGAHLSLKGSYEVLDGLDV  
YGKAGVALVRSDYKFYEDANGTRDHKKGRHTARASGLFAVGAEYAVLPELAVRLEY  
QWLTRVGKYRPQDKPNTAINYNPWIGSINAGISYRFGQGAA

>OMP51\_HAEIN#29#219  
YAGVKAGQASFHDLRALAREYKVGYHRNSFTYGVFGGYQILNQNNLGLAVELGYDDFGRAKGREKGKTVVKHTNHGTHLSLKGSYEVLEGLDV  
YGKAGVALVRSDYKLYNENSSTLKKLGEHHRARASGLFAVGAEYAVLPELAVRLEY  
QWLTRVGKYRPQDKPNTALNYPWIGSINAGISYRFGQGAA

>OMP53\_HAEIN#29#225  
YAGVKAGQGSFHDGINNNGAIAKKGLSSSNYGYRRNTFTYGVFGGYQILNQDNFGLAAELGYDDFGRAKLREAGKPKAKHTNHGAYLSLKGSYEV  
LDGLDVYKAGVALVRSDYKFYEDANGTRDHKKGRHTARASGLFAVGAEYAVLPEL  
AVRLEYQWLTRVGKYRPQDKPNTAINYNPWIGCINAGISYRFGQGEA

>Q65T83\_MANSM#29#213  
YAGARAGWASFHGDVDAFHNSDGLSAKKNSVTYGVFGGYQILNQNNFGLAVELGYDDYGRIRLIEFDAKRGKFTNHGVNLSIKPSYEVLDGLDV  
YARVGAALIRTDYKDYSDAAGHSLKVSPFTAAGLEYALPILPELAMRLEYQWIEN  
VGRDQKWSGYNDADFTPDIGAVTFGLSYRFGQGAA

>Q6XAX4\_MANGL#27#228  
YAGAKAGWASFHDLTQFDHKDGGFEGGINRNSVTYGVFGGYQILNQNNFGLATELGYYDFGRVRGNKTFEGENSDKRAAKHSAHGAHLSLKPSY  
EVVFNLDVYKGVGVALVRNDYYVQQNVAKDSRIKAHNLKPSLLLGALEYAITPEL  
AARVEYQYLNVRVGNLDKAARKTANIEGTNFQYSPDIHSVSAGLSYRFGQGAA

>Q6XAY4\_PASHA#27#239  
YAGAKAGWASFHDLTQFDNTYDYGSGINFGINRNSVTYGVFGGYQILNQNNFGLATELGYYDFGRVRGNNTTFSGDDKEKTAAKHSAHGA  
HLSLKPSYEVVFNLDVYKGVGVALVRNDYYVQQHVDKDSRIKVHNLKPSLLLGALE  
EYAITPELAARVEYQYLNVRVGNLDKAVRTTVGYLIPRGTFNFQYSPDIHSVSAGLSYRFGQGAA

>Q6PSJ7\_PASTR#28#225  
YVGAKAGWATFHHNVNQFNSKYALDERYAGGPTAYGINRNSVTYGVFGGYQIIDNLAVELGYDYFGRVRGNVQVDEQDSKAVKHTAHGTHLSLK  
PSYEVVSGLDVYKGVGAALVRNDYKNYVDQEKSHNLKTSLLLGALEYAITPALAA  
RLEYQWLSNVGNFNKAESKENRLRNVTYRPDIHSVSAGLVYRFGQGAA

>Q6PSK5\_PASTR#28#224  
YVGAKAGWATFHHGVNQFNSKYANDERYSDGPTAYGINRNSVTYGVFGGYQIIDNLAVELGYDYFGRVRGNVGDRAAFKDKHGAHISLKPSYE  
VVSGLDLYGKVGAAALVRNDYKDYANGTKDKSHNLKTSLLGAGVEYAITPALAAR  
LEYQWLSRAGNYGKAEAKAGNVTDLRYSPDAHSVSAGLTYRFGQGAA

>Q3EF13\_ACTSC#29#219  
YAGARVGSAMHHGVDRIDAQFVTDGGLNRNSVTYGVFGGYQILNQNNFGLAAELGYDFFGKTKADAAGYHAAHGASLALKPSYEIYPDLDVYG  
KVGVALVRNMYKAETLSASGDAEKFNKTKASLILGAGVEYAILPELAARLEYQWLS  
KAGNLDSALEADAGYNGTNARYSPDIHSVTAGLSYRFGQGAA

>Q6PSK4\_PASTR#28#221  
YVGAKAGWASFHGDGSIQIDHKNGGKYGINKNSVTYGAFFVGYQIIDNLAAEVGYEYFGRVRGLEQTKAGGSKKTFRHSAGHTTIALKGNYEVISG  
LDTYAKAGIALVNNSYKTVNVDTKVPTKTSRFQSSLILGAGVEYAITPSLGARIEY  
QWLNNAKASYATLRRMGVEGSDYRPDISSVSAGLTYRFGQGAA

>Q5DYH6\_VIBF1#21#185  
YLGQGVGVSHFLGACSSNAIECKNYVTGGGLYGGYQFNSWFALEGSWHDYGSPKTFYGVGDGYYSNATGVDSL SVKLSLPVTDNLDLYAKGGAA  
NYLSVSGNDNVHLGMFESDSSIDDIWEIGA EYALAPNWSLRFGTSIIDGIGNAKTG  
KSDLYFTSLGLTYKF

>Q8E832\_SHEON#33#191  
YAGAGLGQGHYSNGSNPQSYDSVRDRFAGSVYLG YQVNPYLAPELSYQFLGSAYANYEQQISGDFQQVVLAARFGYPLTTSLYPYVKVGAGW  
FGDSEGLRSGSERGFSPIVAAGVEYAFTPRLSGRLEHQYTDLSGADSIGYTDHHLT  
TLGLSWRFG

>P96773\_HAEDU#126#257  
KGD LNRQFKHTAHGANLSLKPSYELPNLDVYKVGMLVRNDYKFI LD LQSKDQHVKDETKLHTLKPSLLLGTGVEYAITPELALRTEYQYLN  
KAGNLYKA AKYKNISTSLAYAPDIHSVSVGLSYRFGQG

>Q87RL4\_VIBPA#23#171  
YVGKMGKSWLEDACVAGQSCDKDGSTLGAFVGYEFNDYIALEAGFDNVGDFDQTSFGGHVEAITLAPKFSLPITQDIALYKGVGGAYVMFDGK  
DDYSYLGAAAGVEFNLSQNV TARAEYQTITDISNNVSKAAGNSATLGVSFKFGGND

>O31154\_VIBCH#24#173  
YVGKVGKSWLDDACLAGQSCEDDDQVVGAF LGYQANKWLSLEAGYDYL GKFTAAGLNDEKVQAVTLAPKLSIPLTEGIALYKGVGGAYVDYGS  
KDDYSYLGAAGLEFNTNHNVTMRLEYQNLTDINNDIVRARAETATLG IAYKFGGSE

>Q7MDC9\_VIBVY#30#182  
YFGGKIGYVNLDDACYVSSPCDDDSAGFLYGYNF SKHISAELGIDLLGEHETNFSNGYFSEAKLAAYTLAPKFSLP LNEKLDFAKIGAA  
IYGDDKDLVPTGSLGLEALTESFKARA EYQRYQDMSDDIVKMDANFFGIGITYL  
FGG

>Q87GX9\_VIBPA#27#181  
YIGGKVGMTTLDDACYLNSPCDDAFAAGLHIGYDFTDFIGLEYGVDFLGDYKANFKSGASTVDTIDGNLTALT LAPKFNWHLNDSWNLF AKIG  
GAYMMSDEKDFVATGSLGAEYSIDRNWSVRAEYQRYQDMSDDVWDDMDANFFGIG  
VNYKF

>Q87JK1\_VIBPA#26#179  
YIGGKVGYNLSNDACYLNEPCDDDSFAAGMHIGYNFNEYVAAEYGVVDYLGDFDTANFKKPGLNTVDGDLWALTLAPKFNLPITDSWNLFQKVGAA  
YMMAGDEKDFVPTGSLGAEYKINYNWSLRAEYQRYQDMSDDIVDDMSDFFGIGFN  
YKFG

>Q44GL1\_CHRSL#62#200  
DVDDDDNTFKGFVGYNFNRYFATEAFYSYDLGRVKLKNGTANTDLESEAYGVSLVGKLPITQWFELFAKAGMAKWETDIDGNLGGASTDLEDND  
GVDPVYGAGAQNFKPFLVRAEYERYDFDSYQIDSFTASVWQWQF

>Q5FA01\_NEIG1#20#175  
EGASGFYVQADAAHAKASSSLGSAKGFSPRISAGYRINDLRFVYTRYKQAPSTDFKLYSIGASVIYDFDTQSPVKPYFGARLSLNRASA  
HLGSDSFSKTSAGLGLVAGVSYAVTPNVDLDAGYRYNYVGKVNKNVRSGELSA  
GVRVKF

>Q3EI33\_ACTSC#24#191  
YVEGDAGYSKLRTSGGLSSKISDGSFSPSVAVGYKVNDWRFALDYTYYGKDEGFGDLGGSGENEVKAYGFGLAAYYDFDLGTSLTPLYGARIS  
ANHIKVNDSFNNASDYFHFSSNTEFGYGAIAVSYNFAPQWDLNVAAEYNRLGEV  
RDVKVNQYGAKVGVRYTF

>Q3QBC3\_9GAMM#28#162  
AKGGYQWASDDSYKHSNPKGAIFGVSSGLQFSPAWSWDVGYYHDDLKADATSVNVKAGLIESALRYDWYLQDNLSVYGRGLGVAYWDMKTHLS  
SKLDATGLSPLGEVGVNYNFTPNVRLSAGYQYIDSIGESN

>Q3QLJ4\_9GAMM#38#150  
ASGDGEGSSDGVGFAYGGYNFNEWFLEANLFASGDLGEDGVVDGAGALTFTPKFTYHFNDTFSFAKVGVASMAVVVDYGPFDADYTGWGLT  
WGVGVNAALTEKLNRLSY

>Q3GMA6\_9GAMM#36#179  
YVGAKVGQFDLDVDGADDPYAGVYAGYNFDPNFGMEAEFVGSDDADYRGGDIDAKTYGAYGTYRYAFPNTGLYAKGKLGVARTEIEATGTSPA  
GFYNGSNKDTSLAGGVGLGYSVNPFSVEAEYDKLGSADLMTVGAQLKF

>Q57RC6\_SALCH#19#171  
  
GTAVAATSTVTGGYAQSDAQGVANKMSGFNLKYRYEQDDNPLGVIGSFYTYTEKDRTNAGDYNKGQYYGITAGPAYRLNDWASIVGVGVGYGK  
FQTTDYPTYKHDTSDYGFYSYGAGLQFNPMENVALDFSIEQSRIRSVVDVGTWIAVG  
YRF

>OMPX\_ENTCL#19#172  
GTAVAATSTVTGGYAQSDMQGVNMKNTNGFNLKYRYEQDNNPLGVIGSFYTYTEKDRTENGSYNKGQYYGITAGPAYRLNDWASIVGVGVGYGKF  
QQTENQGLNRTASNSDYGFYSYGAGMQFNPIENVALDFSIEQSRIRSVVDVGTWIAVG  
GYRF

>Q83ZR1\_ENTAE#25#171  
ATSTVTGGYAQSDMQGKANKAGGFNLKYRYEQDNNPLGVIGSFYTYTEKDNTSNGTYNKGQYYGITAGPAYRLNDWASIVGVGVGYGKFQNNNF  
PNHKSDDSDYGFYSYGAGMQFNPIENVALDFSIEQSRIRSVVDVGTWIAVG  
YRF

>Q669E5\_YERPS#18#174  
  
TAGSAFAGQSTVSGGYAQSDYQGVANKSSGFNLKYRYEWSDSQLGYITSFTHTEKSGFGDEAVYNKAQYNAITGGPAYRINDWASIVGLVGVGH  
GRFTQNESAFVGDHSTSDYGFYAGLQFNPAENVALDVSIEQSRIRSVVDVGTWV  
AGVGYTF

>Q54470\_SERMA#27#172  
STVSAGYAQGDQGVANKADGFNLKYRYEFDNNPLGVIGSFTHLEKDGSDGFYNKAQYNSISAGPAYRINDWASIVGLVGLGYGKFTTNAQNG  
TSRHDTADYGFTYAGLQFNPIENVALDVGYEQNRIRSVVDVGTWNVGVGYRF

>Q5J664\_SERMA#27#172  
  
STVSAGYAQGDQGVANKADGFNLKYRYEFDNNPLGVIGSFTHLEKNRSENGFYKKSQYDSITAGPAYRFNDWASIVGVGVGYGKNIDNAQAG  
GNKGGNSDYGFYAGLQFNPIENVALDVGYEQSRIRSVVDVGSWNVGVGYRF

>Q6D3I0\_ERWCT#17#171  
  
GAGSAFAGQSTVTAGYAQGDAQGVQNKVKGFNLKYRYEQDNNPLGVISSFTYLEKNGSDDGFYSKGQYMGFTAGPAYRLNDWASIVGVGVGSHG  
KVTSNNTNGQDNASNDYGFYAGVQFNPIQDVALDVGYEQSRIRSVVDVGAWVG  
VGYRF

>Q7N474\_PHOLL#26#175  
STISGGYAQSHVKFNGEKIDGNPKGKGFNLKYRYELDDNNGFVVISLTYTHKGDFKFDVNYYSLMAGPAYRFNDYVSAYGMIGAARGEFEVPVLN  
GSYTEGKSSVSYGLGLQFNPNVNTIDTSYEFYAKFNARLDGLKVGTVVGVGYRF

>AIL\_YEREN#32#178  
GYAQSHVKENGYTLDDNPKGFNLKYRYELDDNNGFVVISLTYTHQGYDFYGSNKGHGDVDYYSVTMGPSFRINEYVSLYGLLGAHGVKASV  
FDESIASAKTSMAYGAGVQFNPLPNFVIDASYEYSKLDISKVGTWMLGAGYRF

>Q7N473\_PHOLL#27#181  
  
TISGGYAQSHVKVDGEKLDENPKGFNLKYRYELDDNNGFVVISLTYTHQGYDYGGSSKIGTGDLDYSLMAGPTYRFNEYVSVYGLVGAHRKV  
KGEFSAYGYDFEYKESKTSAAAGLGLQFNPNVNVVIDASYEYSKLDGDFKVGTVWVIG  
AGYRF

>Q66AL7\_YERPS#27#179

ASNTVSFGYAQSTLKDGEKIGKDNKGFNLYRHELDVSLGIVASFTHTKQNYGMPGDSGKRKVEYYSLMVGPSWRFNEFVSAYALIGATQGK  
STHTKPRMVSNTVSKTSMGYGAGLQFNVPKHVAIDTAYEYAKIEDVKIGTWIVGVG  
YRF

>Q8Z920\_SALTY#29#174  
SFGYAQTHLSSLKNSDSKDLRGFNFKYRYEFNETWGMLGSFTATRNMENYTWKEGKLHKNGSDSDVYGSLMFGPTYRFNDYVSLYGNAGIATM  
KFNKHSKEESFAYGAGVIFNPVKISIDASWEASRFFAVDTNTFGVSVGYR

>Q6DB56\_ERWCT#23#182  
ATQTVSLGYAQAKVEDFKDLKGVTAHYHQGDSALGIIIGSFTYLGGKKDYNQSSIGFSSHESDDVKYYSLMVGPSYQINDVVSIIYGLAGLGHAK  
VKNSYNYTLPNQSGSESYSESKTKFAYGAGIQITPVANWAIDIGYEGTKIYEARVD  
GFNIGVGYRF

>Q04817\_SALTY#29#185  
VSVGYAQSRIEHFKDIRGVNLKYRYEAQTPLGLMASFSWQSGKRGESGGIPGGMSWRDDVKATYWSLMAGPAVRVNLVSLYALAGAGTGRAEV  
KERISMPGYNGRFTGSERTGFAWGAGVQFNPVENVIDLGYEGSKVGAAKLNGVN  
VGVGYRF

>PAGC\_SALTY#31#185  
GYAQSKVQDFKNIRGVNLKYRYEDDSPVSFISLSYLYGDRQASGSVEPEGIHYHDKFEVKYGLMVGPAAYRLSDNFSLYALAGVGTVKATFKE  
HSTQDGDGFSNKKISSRKTGFAWGAGVQMNPLENIVVDVGYEGSNISSTKINGFNVG  
VGYRF

>Q8XBY9\_ECO57#13#177  
SAGYAQSKVQDFKNIKGVNLQYRYEWDSPVSVVGSFSYMKGDWADSHRDEADDFYRHQADIKYYSFLAGPAYRLNDYISFYGLVGISHTKAKGD  
YEWNRNSVGADESGLSESVSKKSTDFAYAAGVI INPWGNMSVNVGYEGTKADIYG  
KHSVNGFTVGVGYRF

>Q9XD46\_VITS1#32#191  
TVSIGYSHGKISGADKLVGTAKYNYQFDQQPWGVMNTLYTMGGKQRNNITSSNQLYENDVDVDYYSAGVGPSYRINPNVNVYGVGVAKADTK  
GTQKQNRTRGRIYNIDNIDTNVMYGAGVQYNPAPNWSVDVGYESSRVNDGYDKRSMN  
AFNVGVGYRF

>Q8FEX5\_ECOL6#32#199  
SAGYLQTHDMPGSDDLKGINVKYRYEFTDTLGLVTSFSYANAKDEQKTHYSSTRWHEDSVRNRWF SMMAGPSVRVNEWFSAYAMAGMAYSRVS  
TFSGDYLRLVTDNKGKTHDVLGTGSDDGRHSNTSLAWGAGVQFNPTESVAIDLAYEGS  
GSGDWRTDGFIVGVGYKF

>Q6H9T1\_9CAUD#19#199  
GAPAWASEHQSTLSAEYLHASTNVSGSDDLNGINVKYRYEFTDTLGMVTSFSYAGDKNRQLTHYSSTRWHEDSVRNRWFSVMAGPSVRVNEWFS  
AYAMAGVAYSRSVTSFGDYLRLVTDNKGKTHDVLGTGSDDGRHSNTSLAWGAGVQFNPT  
TESVAIDIAIEGSGSGDWRTDGFIVGVGYKF

>Q7DBL7\_ECO57#29#199  
STLSAGYLHASTNVPGSDDLNGINVKYRYEFMDALGLITSFSYANAEDQKTRYSDTRWHEDSVRNRWFSVMAGPSVRVNEWFSAYAMAGVAYS  
RVSTFSGDYLRLVTDNKGKTHDVLGTGSDDGRHSNTSLAWGAGVQFNXPESVAIDIAIEGSGSGDWRTDGFIVGVGYKF

>Q8X428\_ECO57#7#108  
SVMAGPSVRVNEWFSAYAMAGVAYSRSVTSFXGDYLRLVTDNKGKTHDVLGTGSDDGRHSNTSLAWGAGVQFNXPESVAIDIAIEGSGSGDWRTDGF  
IVGVGYKF  
>O33796\_SALTY#23#153

AGYKNTVSIYAYTDLGSLGNSGNANGANIKYNWEDLDSGFGAMGSVITYTSADVNNYGYKVGADADYTSLLVGPSYRFNDYLNAYVMIGAANGHIK  
DNWGNSDNKTAFAYGAGIQLNPVENIAVNASYEHTSF

>Q7N5I2\_PHOLL#43#182  
LNGATLSYRYELDDQWGLLSSFTFAKGDEKETKYTGEDLKYYSVMIGPTYRINDYISLYGQLGLSRINDKSTAHYSGGYIEKESTSKNTLWG  
AGFIINPTTNTSITAGYEGSRFSIKDGNKDHLSSTNGFNITVGYRF

>AIL\_YERPS#91#182  
DGFYKYYSVTAGPVFRINEYVSLYGLLGAGHGKAKVSSIFGQSESRSKTSLAYGAGLQFNPHPNFVIDASYEYSKLDLVKVGTMWLGAGYRF

>Q66BP0\_YERPS#98#183  
YYSLMVGPSVHFNEFFSMYALLGIGHGKAKASVLGYGKKEEQDSLAYGVGMQFNPLNNIAIDASYEYAKLKNANIGTWVLGIGYRF

>Q57K85\_SALCH#72#178  
GSQRYSDSSNGRVTTRYSSLLAGPSWKINNQLSLYSQVGPVLLHQRDHGINESDSKVGYGYSAGVAYTPVSNVAITLGYEGADFDATHNSGSL  
NSNGFNLGVGYRF

>Q8ZQG1\_SALTY#92#172  
GPSYRFNEYLNAYVMAGLGHGHIDDKRDNSGKKTGFAYGAGVQINPVENIAINASYEYSRFSAYDSKVNAGTWVLGVGYSF



## 8.2 Sequence alignments used for homology model generation

Sequence alignments in the .pir format used by modeller 9.11

### OmpA

```
>P1;1BXW
structureX:1BXW:1 :A:171 :A:OMPA:E COLI::
-APKDNTWYTGAKLGWSQYHDTGLINNGPTHENKLGAGAFGGYQVNPYVG--FEMGYDWLGRMPYKGSVENGAY
KAQGVQLTAKLGY-P-I-T-D-D-LDIYTRLGGMVWRAD-TYSNVYGNHDTGVSPV-FAGGVEYA-I-T-P-EI
ATRLEYQWTNN-IG-DAH-TIGT-RP-DNG--MLSLGVSYRFG*

>P1;0422
0422:1: : :182 : :OMPA:B AFZELII::
--QSKSKTMVEDDFDFEKLLEK---EESVR----RLFGIGFGIGY-PLTNIIISVPYVDIDLGYGGFVGLKPNN
FMPYVVMGIDLLFKDEIHKNTMISGGIGIGADWSKGSPEKSNENLEGDVNEQQTSLENRIGVVIRLPLVIEYSF
LKNIVIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI*
```

### OmpW

```
2F1T (OmpW) from E. coli
-EAGEFFMRAGSATVRPTEGGGFSVTNNTQLGLTFTYMATDNIGVELLAATPFRHKIGTRATGDIATVHHLPT
LMAQWYFGD-AS-SK-FRPYVGAGINYYTFFDNGFNDHGKEAGLSDSLKDSWGAAGQVGVDY-LINR-DWLVNM
SVWYMDID-TTANYKLGAQQHDSVRLDPWVFMSAGYRFH*

>P1;0422
0422:1: : :182 : :OMPA:B AFZELII::
QSKSKTMVEDDFDFEKLLEK-EESVRRLFGIGFGIGY--P-LTNIIISVPYVDIDLGYGGFVGLKPNNFMPYVV
MGIDLLFKDEIHKNTMISGGIGIGADWSKGSPEKSNENLEGD-VNEDQQ-TSLENRIGVVIRLPLVIEYSFLKNI
VIGFKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI--*
```

### OmpX

```
>P1;1QJ8
structureX:1QJ8:1 :A:148 :A:OMPX:E COLI::
-----ATSTVTGGYAQSDAQGQ-MNKMGGFNLKYRYEEDNSPLGVIGSFYTEKSRTASSGDY--NKNQYYGITA
GPAY-R-I-N-D-WA-SIYGVGV-GYGK-FQTTEYPTYKNDT-SDYGFS---YGAGLQFNPMENVALDFS
Y-----E-Q-S-R-I-R-S-V-D--VGTWIAGVGYRF-*

>P1;0422
0422:1: : :182 : :OMPA:B AFZELII::
QSKSKTMVEDDFDFEKLLEKEESVRRLFGIGFGIGY-PLTN-IIISVPYVDIDLGYGGFVGLKPNNFMPYVVMGI
DLLF--KDEIHKNTMISGGIGIGADWSKGSPEKSNENLEGDVNEQQTSLENRIGVVIRLPLVIEYSFLKNIVIG
FKAVATIGTTMLFGNPMSFEGARFNFLGTGFIKIYI*
```

## 8.3 Conditions for membrane protein crystallisation screen

Screen used: Memb-PASS screen from JENA Bioscience.

Number	Precipitant	Buffer	pH	Additive 1	Additive 2
A 1	48 % v/v PEG 400	100 mM Sodium Citrate	4.5	210 mM Sodium Acetate	1.12 % v/v Glycerol
A 2	32 % v/v PEG 400	100 mM Sodium Acetate	5.5	120 mM Potassium Chloride	None
A 3	46 % v/v PEG 400	100 mM Sodium Acetate	5.5	230 mM Sodium Malonate	None
A 4	42 % v/v PEG 400	100 mM ADA	6.5	290 mM Magnesium Acetate	None
A 5	45 % v/v PEG 400	100 mM HEPES	7.5	180 mM Sodium Acetate	450 mM TMAO
A 6	28 % v/v PEG 400	100 mM Tris HCl	8.5	290 mM Lithium Nitrate	None
A 7	36 % v/v PEG 400	100 mM CAPSO	9.5	170 mM Magnesium Chloride	None
A 8	36 % v/v PEG 400	None		120 mM Magnesium Sulfate	3.67 % v/v Glycerol
A 9	47 % v/v PEG 400	None		170 mM Ammonium Sulfate	None
A 10	41 % v/v PEG 400	None		220 mM Sodium Malonate	None
A 11	31 % v/v PEG 400	None		40 mM Sodium Sulfate	None
A 12	32 % v/v PEG 400	None		240 mM Sodium Sulfate	None
B 1	30 % v/v PEG 550 MME	100 mM Sodium Citrate	4.5	None	3.22 % v/v Glycerol
B 2	37 % v/v PEG 550 MME	100 mM Sodium Acetate	5.5	None	None
B 3	25 % v/v PEG 550 MME	100 mM Sodium Acetate	5.5	200 mM Calcium Chloride	None
B 4	33 % v/v PEG 550 MME	100 mM ADA	6.5	170 mM Magnesium Sulfate	None
B 5	38 % v/v PEG 550 MME	100 mM ADA	6.5	140 mM Sodium Bromide	None
B 6	39 % v/v PEG 550 MME	100 mM HEPES	7.5	100 mM Calcium Chloride	None
B 7	34 % v/v PEG 550 MME	100 mM Tris HCl	8.5	220 mM Ammonium Nitrate	None
B 8	23 % v/v PEG 550 MME	100 mM CAPSO	9.5	60 mM Lithium Chloride	None
B 9	29 % v/v PEG 550 MME	None		290 mM di Potassium hydrogen Phosphate	None
B 10	23 % v/v PEG 550 MME	None		280 mM Sodium Bromide	None
B 11	22 % v/v PEG 550 MME	None		None	None
B 12	30 % v/v PEG 550 MME	None		200 mM Calcium Chloride	None
Number	Precipitant	Buffer	pH	Additive 1	Additive 2
C 1	32 % v/v PEG 550 MME	None		220 mM Sodium Malonate	None
C 2	38 % v/v PEG 550 MME	None		160 mM Magnesium Acetate	None
C 3	33 % v/v PEG 550 MME	None		250 mM Sodium Acetate	None
C 4	25 % w/v PEG 1000	100 mM Sodium Citrate	4.5	190 mM Potassium Chloride	None
C 5	35 % w/v PEG 1000	100 mM Sodium Citrate	4.5	190 mM Ammonium Nitrate	None
C 6	27 % w/v PEG 1000	100 mM Sodium Acetate	5.5	140 mM Ammonium Nitrate	None
C 7	34 % w/v PEG 1000	100 mM ADA	6.5	None	None
C 8	20 % w/v PEG 1000	100 mM HEPES	7.5	200 mM Ammonium Nitrate	None
C 9	28 % w/v PEG 1000	100 mM HEPES	7.5	100 mM Sodium Malonate	None
C 10	22 % w/v PEG 1000	100 mM Tris HCl	8.5	None	None
C 11	33 % w/v PEG 1000	100 mM Tris HCl	8.5	None	None
C 12	35 % w/v PEG 1000	100 mM CAPSO	9.5	180 mM Potassium Chloride	None
D 1	28 % w/v PEG 1000	100 mM CAPSO	9.5	290 mM Lithium Chloride	None
D 2	19 % w/v PEG 1000	None		240 mM Lithium Nitrate	None
D 3	33 % w/v PEG 1000	None		270 mM Ammonium Sulfate	None
D 4	29 % w/v PEG 1000	None		210 mM Sodium Bromide	3.80 % v/v Glycerol
D 5	32 % w/v PEG 1000	None		None	2.35 % v/v Glycerol
D 6	35 % w/v PEG 1000	None		180 mM Potassium Chloride	4.16 % v/v Glycerol
D 7	36 % w/v PEG 1000	None		20 mM Zinc Acetate	None
D 8	None	None		2 M Ammonium Sulfate	None
D 9	34 % w/v PEG 2000 MME	100 mM Sodium Citrate	4.5	None	540 mM TMAO
D 10	28 % w/v PEG 2000 MME	100 mM Sodium Citrate	4.5	None	None
D 11	26 % w/v PEG 2000 MME	100 mM Sodium Citrate	4.5	20 mM Zinc Acetate	None
D 12	23 % w/v PEG 2000 MME	100 mM Sodium Acetate	5.5	100 mM Sodium Acetate	None

Number	Precipitant	Buffer	pH	Additive 1	Additive 2
E 1	29 % w/v PEG 2000 MME	100 mM Sodium Acetate	5.5	150 mM Sodium Nitrate	None
E 2	28 % w/v PEG 2000 MME	100 mM Sodium Acetate	5.5	None	None
E 3	26 % w/v PEG 2000 MME	100 mM ADA	6.5	130 mM Sodium Sulfate	None
E 4	18 % w/v PEG 2000 MME	100 mM ADA	6.5	None	None
E 5	29 % w/v PEG 2000 MME	100 mM HEPES	7.5	290 mM Lithium Chloride	None
E 6	17 % w/v PEG 2000 MME	100 mM Tris HCl	8.5	280 mM Lithium Nitrate	None
E 7	34 % w/v PEG 2000 MME	100 mM ADA	6.5	20 mM Zinc Acetate	None
E 8	29 % w/v PEG 2000 MME	100 mM CAPSO	9.5	None	None
E 9	17 % w/v PEG 2000 MME	100 mM CAPSO	9.5	120 mM Lithium Sulfate	None
E 10	18 % w/v PEG 2000 MME	None		270 mM Magnesium Sulfate	None
E 11	22 % w/v PEG 2000 MME	None		200 mM Magnesium Chloride	None
E 12	31 % w/v PEG 2000 MME	None		210 mM Ammonium Sulfate	None
F 1	29 % w/v PEG 2000 MME	None		170 mM Magnesium Chloride	None
F 2	26 % w/v PEG 2000 MME	None		110 mM Calcium Chloride	None
F 3	27 % w/v PEG 2000 MME	None		160 mM Magnesium Chloride	None
F 4	30 % w/v PEG 2000 MME	None		230 mM Sodium Acetate	None
F 5	27 % w/v PEG 2000 MME	None		130 mM Sodium Acetate	None
F 6	33 % w/v PEG 2000 MME	None		240 mM Magnesium Chloride	None
F 7	16 % w/v PEG 4000	100 mM Sodium Citrate	4.5	None	None
F 8	14 % w/v PEG 4000	100 mM Sodium Citrate	4.5	210 mM Magnesium Chloride	None
F 9	16 % w/v PEG 4000	100 mM Sodium Acetate	5.5	120 mM Ammonium Nitrate	None
F 10	20 % w/v PEG 4000	100 mM Sodium Acetate	5.5	None	None
F 11	16 % w/v PEG 4000	100 mM Sodium Acetate	5.5	None	870 mM TMAO
F 12	20 % w/v PEG 4000	100 mM Sodium Acetate	5.5	200 mM Magnesium Chloride	None
Number	Precipitant	Buffer	pH	Additive 1	Additive 2
G 1	13 % w/v PEG 4000	100 mM ADA	6.5	290 mM Sodium Acetate	None
G 2	19 % w/v PEG 4000	100 mM ADA	6.5	50 mM Lithium Nitrate	None
G 3	16 % w/v PEG 4000	100 mM ADA	6.5	20 mM Zinc Acetate	None
G 4	17 % w/v PEG 4000	100 mM HEPES	7.5	270 mM Ammonium Nitrate	None
G 5	13 % w/v PEG 4000	100 mM HEPES	7.5	150 mM Ammonium Nitrate	420 mM TMAO
G 6	13 % w/v PEG 4000	100 mM Tris HCl	8.5	110 mM Lithium Sulfate	1.67 % v/v Glycerol
G 7	12 % w/v PEG 4000	100 mM Tris HCl	8.5	240 mM Lithium Acetate	None
G 8	21 % w/v PEG 4000	100 mM Tris HCl	8.5	None	None
G 9	18 % w/v PEG 4000	100 mM Tris HCl	8.5	270 mM di Potassium hydrogen Phosphate	None
G 10	14 % w/v PEG 4000	100 mM CAPSO	9.5	200 mM Sodium Sulfate	None
G 11	20 % w/v PEG 4000	100 mM CAPSO	9.5	None	None
G 12	23 % w/v PEG 4000	None		280 mM Sodium Bromide	None
H 1	13 % w/v PEG 4000	None		210 mM Magnesium Chloride	1.37 % v/v Glycerol
H 2	20 % w/v PEG 4000	None		230 mM Potassium Thiocyanate	1.11 % v/v Glycerol
H 3	18 % w/v PEG 5000MME	100 mM Sodium Citrate	4.5	220 mM Magnesium Sulfate	None
H 4	17 % w/v PEG 5000MME	100 mM Sodium Acetate	5.5	250 mM Lithium Nitrate	4.57 % v/v Glycerol
H 5	18 % w/v PEG 5000MME	100 mM Sodium Acetate	5.5	220 mM Sodium Sulfate	470 mM TMAO
H 6	20 % w/v PEG 5000MME	100 mM ADA	6.5	70 mM Lithium Sulfate	None
H 7	12 % w/v PEG 5000MME	100 mM HEPES	7.5	180 mM Ammonium Sulfate	None
H 8	20 % w/v PEG 5000MME	100 mM Tris HCl	8.5	200 mM Sodium Sulfate	None
H 9	15 % w/v PEG 5000MME	100 mM CAPSO	9.5	150 mM Lithium Nitrate	None
H 10	13 % w/v PEG 5000MME	None		50 mM Calcium Chloride	None
H 11	14 % w/v PEG 5000MME	None		20 mM Zinc Acetate	None
H 12	16 % w/v PEG 5000MME	None		None	None